

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 5, 2004, 17:59:26 ; Search time 4223 Seconds
(without alignments)
5614.170 Million cell updates/sec

Title: US-09-943-857-4

Perfect score: 2864

Sequence: 1 SMNSRCPAGRLGSLVPTAKLA.....DNFRTAGYDALMTNPSFFV 547

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -MODEL=frame+ p2n.model -DEV=xlh

-Q=/cgn2_1/USPTO spoof/US09943857/runat 29072004_090900_29732/app query.fasta_1.711
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPECL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfo -NORM=ext -HEAPSZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09943857@cgn 1.1.2568 @runat 29072004_090900_29732 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2671	93.3	1855	8 CCLIP3	X66006 C.cylindrac
2	2600	90.8	1647	8 AY464467	AY464467 Candida c
3	2513	87.7	1647	6 A81171	A81171 Sequence 1
4	2513	87.7	1688	12 AF044078	AF044078 Synthetic
5	2416	84.4	1857	8 CCLIPASE	X16712 C. cylindra
6	2415	84.3	1733	6 A48377	A48377 Sequence 6
7	2415	84.3	1733	8 CCLIP1	X64703 C.cylindrac
8	2393	83.6	1605	6 A81172	A81172 Sequence 2
9	2392	83.5	1532	6 AX709921	AX709921 Sequence
10	2347	81.9	1812	8 CCLIP5	X66008 C.cylindrac
11	2268	79.2	1785	8 CCLIP4	X66007 C.cylindrac
12	2212	77.2	2043	8 CCLIP2	X64704 C.cylindrac
13	2034	71.0	1541	6 AX709925	AX709925 Sequence
14	2020	70.5	1548	6 AX709923	AX709923 Sequence
15	1903	66.4	1511	6 AX709927	AX709927 Sequence
16	1875	65.5	1469	6 AX709919	AX709919 Sequence
17	1107	38.7	1828	6 A74255	A74255 Sequence 11
18	1095	38.2	1635	8 GCU02525	U02525 Geotrichum
19	1092	38.1	1692	6 E02678	E02678 cDNA encodi
20	1089	38.0	1635	6 A48376	A48376 Sequence 5
21	1089	38.0	1635	8 GCU02387	U02387 Geotrichum
22	1088	38.0	1635	8 GCU02622	U02622 Geotrichum
23	1087	38.0	1783	8 AB000260	AB000260 Geotrichu
24	1084	37.8	1635	8 GCU02524	U02524 Geotrichum
25	1081	37.7	1635	8 GCTAGL	X81656 G.candidum
26	1081	37.7	1635	8 GCU02625	U02625 Geotrichum
27	1081	37.7	1692	8 GGECLR	X78032 G.geotrichu
28	1078	37.6	1674	6 E02497	E02497 cDNA encodi
29	1078	37.6	1767	8 GCALIP2	D00697 Geotrichum
30	1070	37.4	1635	8 GCU02623	U02623 Geotrichum
31	1069	37.3	1635	8 GCU02541	U02541 Geotrichum
32	938.5	32.8	2045	6 A45569	A45569 Sequence 7
33	938.5	32.8	2045	6 A47607	A47607 Sequence 7
34	938.5	32.8	2045	6 AR019333	AR019333 Sequence
35	801	28.0	4295	8 AY250996	AY250996 Aspergill
36	693	24.2	1934	8 AK107259	AK107259 Oryza sat
37	632	22.1	2000	8 ANGLAA5	X56442 A.niger gla
38	536.5	18.7	13546	1 AE011751	AE011751 Xanthomon
39	495	17.3	1971	5 GGA306928	AJ306928 Gallus ga
40	491.5	17.2	2490	5 TCACBR	X03439 Torpedo cal
41	489	17.1	2357	5 TMACHE	X05497 Torpedo mar
42	484	16.9	2033	3 AF080184	AF080184 Meloidogy
43	482	16.8	2133	3 AF075718	AF075718 Meloidogy
44	481	16.8	2089	6 AX306139	AX306139 Sequence
45	481	16.8	2089	10 MMACHE	X56518 Mouse mRNA

ALIGNMENTS

RESULT 1

CCILIP3
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
gene
CDS
sig_peptide
mat_peptide
ORIGIN
Alignment Scores:
Pred. No.:
Score:
Percent Similarity:
Best local Similarity:
Query Match:
DBs:

CCILIP3
C.cylindracea Lip3 gene.
X66006 S55937
X66006.1 GI:296933
lip3 gene; lipase.
Candida cylindracea
Candida cylindracea
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
Lotti,M., Grandori,R., Fusetti,F., Longhi,S., Brocca,S.,
Tranontano,A. and Alberghina,L.
Cloning and analysis of Candida cylindracea lipase sequences
Gene 124 (1), 45-55 (1993)
83178975
840480
2 (bases 1 to 1855)
Longhi,S.
Direct Submission
Submitted (06-MAY-1992) S. Longhi, Universita' degli Studi di
Milano, Via Celoria 26, 20133 Milano, ITALY
See X66006-8, X16712, X64703 & X64704
X66006, X66007 and X66008 are related sequences in the description
of C.cylindracea Lip1 and Lip2 sequences (x64703, x64704)).
Location/Qualifiers
1..1855
/organism="Candida cylindracea"
/mol_type="genomic DNA"
/strain="ATCC 14830"
/db_xref="taxon:44322"
/clone_lib="SACI genomic"
84..1733
/gene="Lip3"
84..1733
/gene="Lip3"
/EC_number="3.1.1.3"
/codon_start=1
/product="triacylglycerol lipase"
/protein_id="CAA46805.1"
/db_xref="GI:1345989"
/db_xref="GOA:P32947"
/db_xref="SWISS-PROT:P32947"
/translation="MKLALLLIASVAAAPTAKLANGDTITGLNAIINEAFIGPFA
EPVGNLRKDPVPVPSGLNGQKFTLYPLCMQONPEGTFEENLGTALDLVMSKVF
QAVLPQSEDCLTINVRPCTKAGANLVMLTIFGGFGIGSPTIPPPAQMYTKSVL
KPTIILHAVNVRVASWFFLAGDDIKAEQSGNAGLKQRLGMQWADNIAGFGDTSKV
TGFIELAGSMVLCHLIWNQDNTYKGPLFRAGIMQLGMVPLQDVCTGYNEIYDL
FVSAGGCSASDKLARSLSLTDLTADNTNPGFLAYSLRLLYLPRPDGKNIIDDM
YKIVRDGKYSASVPVIIGQNDDEGTIFGLLLNVTNAQARAYFKQLFIASDAEIDTL
MAAPVQDITQGLPDTGIFNAITPQKRLILAVGLDLTAHRRYFLNHFGGGTKYSFL
LKQLGLPIMGTHFANDIVQDVLGGSVIYNNAFTAFATDLDPTNLTAGLVNWFKFT
SSLOLGNLMMINALGIYTKCNFRTAGYDALMTNPLIFV"
84..218
/gene="Lip3"
219..1730
/gene="Lip3"
/product="triacylglycerol lipase"
/EC_number="3.1.1.3"

1.01e-214 Length: 1855
2671.00 Matches: 511
95.90% Conservative: 4
95.16% Mismatches: 22
93.26% Indels: 0
8 Gaps: 0

US-09-943-857-4 (1-547) x CCILIP3 (1-1855)
11 LeuGlySerValProThrAlaIysLeuAlaGlyAspThrIleThrGlyLeuAspAla 30


```

QY 331 AspMetTyrIysLeuValArgAspGlyLysTyrAlaSerValProValIleIleGlyAsp 350
Db 997 GACATGATCGCTTGTGCGGAGGCAAGTATGCCAATCCCTGTGATCATCGCGAC 1056
QY 351 GlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnAlaGln 370
Db 1057 CAGAACGACGAGGCGACCTTCTTTGGCACCAGCAGCTTGAACTGACCGATGCCAG 1116
QY 371 AlaArgAlaTyrPheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeu 390
Db 1117 GCCCGCGAGTATTTCAGCAGAGCTTTGTCCAGCCGACGACCGGAGATCGACGTTG 1176
QY 391 MetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsn 410
Db 1177 ATGACGCGCTACCCCGCGACATCATCCAGGCGAGCCGTTGCACACGGGTATTCTCAAC 1236
QY 411 AlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHis 430
Db 1237 GCCTCACCCCGCAGTTCAGGAATCAGCGCGGTGCTCGCGACCTTGGCTTTACGCTT 1296
QY 431 AlaArgTyrPheLeuAsnHisPheGlnGlyThrIysTyrSerPheLeuSerLys 450
Db 1297 GCTGCTGCTACTTCTCTCAACCTACACCGCGCGCACCAAGTACAGTTCCTCAGCAAG 1356
QY 451 GlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAsp 470
Db 1357 CAGCTCAGCGGCTTGCCTGCTCGGAACGTTCCACTCCAAACGACATTTGTTCCAGGAC 1416
QY 471 TyrLeuLeuGlySerGlySerValIleTyrAsnAlaPheIleAlaPheAlaThrAsp 490
Db 1417 TACTTGTGGCAGCGCGCTCGCTCATCTACAAACGCGTTCATCGCTTTCGCGACGAC 1476
QY 491 LeuAspProAsnThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGln 510
Db 1477 TTGGACCCCAACACCGCGGTTGTTGGTGAAGTGCCTGAGTACACAGCAGCACCCAG 1536
QY 511 SerGlyAsnAsnLeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPhe 530
Db 1537 AGCGGCAACAACCTGATGATGATCAACGCGCTTGGGCTTGTACACCGCGAAGACAACCTC 1596
QY 531 ArgThrAlaGlyTyrAspAlaLeuMetThrAsnProSerSerPhePheVal 547
Db 1597 CGACCGCGGCTACGCGGTGTTCTCCAAACCGCGAGCTTCTTTGTG 1647

RESULT 4
AF044078
LOCUS
DEFINITION
  Synthetic construct triacylglycerol hydrolase (lip1) gene, complete cds.
ACCESSION
  AF044078
VERSION
  1
KEYWORDS
  synthetic construct
  synthetic construct
  artificial sequences.
ORGANISM
  Brocca,S., Schmidt-Dannert,C., Lotti,M., Alberghina,L. and Schmidt,R.D.
REFERENCE
  1 (bases 1 to 1688)
  Design, total synthesis and functional overexpression of the Candida rugosa lip1 gene coding for a major industrial lipase
  Protein Sci. (1998) in press
  2 (bases 1 to 1688)
  Brocca,S., Schmidt-Dannert,C., Lotti,M., Alberghina,L. and Schmidt,R.D.
  Direct Submission
  Submitted (22-JAN-1998) Institute of Technical Biochem., University of Stuttgart, Altmandering 31, Stuttgart 70569, Germany
  Location/Qualifiers
  1..1688
  /organism="synthetic construct"
  /mol_type="genomic DNA"
  /db_xref="taxon:32630"
  /note="designed gene encoding lip1 of Candida rugosa"
FEATURES
  source

```

gene
CDS

```

19..1668
/genes="lip1"
19..1668
/genes="lip1"
/EC_number="3.1.1.3"
/function="lipase"
/note="Lip1; codon optimized for expression in Saccharomyces cerevisiae"
/codon_start=1
/product="triacylglycerol hydrolase"
/protein_id="AAC02093.1"
/db_xref="GI:2852390"
/translation="MELALSLIASVAAPATLTIANGDTITGLNAINIEAFGLIPPA
EPPVGNLRKDPVPYSGSDGKFTSYGSPQONPEGTYEENLPKAAIDLVMQSKVF
BAPSSSECLATINVRPPTKAGANLPVLMIPGGFVGGCTSTFPQAMITYKSTAM
GXPLHVSVNYRVSSWGLAGDEIKAGSANGLQKRLQGMQVADINAFAGGDPPTKV
TIFGASGMSVMCHILMNDGNTYKGFPLFRAGINQSGAMVPSDAVDGIGYNEIFDL
LASNCGGASDKLACLRGSSDTLEATNPTFGFLAYSLSLSYLPDPDGVNITDLM
YALVRGKYANIPVITGQNDGCTFFGTSLSNVTTDAQAREFKQSFVHASDAEIDTL
MTAYPGDITQGGPFDITGILNALTPOPKRISAVLGDGLFLARRYFLNHYTGGTKYSEL
SKOLSLPLVLTGTFHSDIYFQDYLLGSSGLIYNNAFIAFATDLDPTAGLLVKNPEYT
SSSQSNLMLMINALGLYTGKNFRTAGDYDALFSPNPSPFV"

```

ORIGIN

Alignment Scores:

Pred. No.:	1..66e-201	Length:	1688
Score:	2513.00	Matches:	470
Percent Similarity:	93.48%	Conservative:	32
Best Local Similarity:	87.52%	Mismatches:	35
Query Match:	87.74%	Indels:	0
DB:	12	Gaps:	0

US-09-943-857-4 (1-547) x AF044078 (1-1688)

QY	11	LeuGlySerValProThrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAla	30
Db	55	TTGTGCTGCGCCCAACCGCCACTTTTGGCTAACGGTGACACCATCACCAGTTTGAACGCC	114
QY	31	IleIleAsnGluAlaPheLeuGlyIleProPheAlaGluProValGlyAsnLeuArg	50
Db	115	ATCATCAAGAAAGCCCTTCTTGGGTATTTCCATTGCGGAACCCACCACTTGGTAACCTTGA	174
QY	51	PheLysAspProValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGly	70
Db	175	TTCAAGGACCCAGTTCCTACTCCGTTCTTGGATGGTCAAAAGTTCATCTTACCGGT	234
QY	71	ProSerCysMetGlnGlnAsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAla	90
Db	235	CCATCTGTATGCAACAAACCCAGAGGTACCTACGAAAGAACTTGCACAAAGGCGAGCT	294
QY	91	LeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCys	110
Db	295	TTAGATCTGTTATGCAATCCAAAGTTTTCGAAGCTGTTTCTCCATCTTCTGAAGACTGT	354
QY	111	LeuThrIleAsnValValArgProGlyThrLysAlaGlyAlaAsnLeuProValMet	130
Db	355	TTGACCATTAAGTTGTAGACCACCGCGCAAGAGCTGGTCCCACTTGCACGATTATG	414
QY	131	LeuTrpIlePheGlyGlyPheGluIleGlySerProThrIlePheProAlaGln	150
Db	415	TTGTGGATCTTTGGTGGTGGTTTGAAGTTGGTGTACTGACCTTCCCTCCAGCCCAA	474
QY	151	MetValThrLysSerValLeuMetGlyLysHisIleIleHisAlaValAsnTyrArg	170
Db	475	ATGATTACCAAGTCTATTGCTATGGTAAAGCAATCATCCACGTTCTTGCTCAACTACAGA	534
QY	171	ValAlaSerTrpGlyPheLeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsnAla	190
Db	535	GTCTCGAGCTGGGGTTTCTTGGCTGTCAGCAAAATCAAGGCCAAGGTTCTGCCAACGCC	594
QY	191	GlyLeuLysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGly	210
Db	595	GGTTTGAAGGACCAAGATTGGGTATGCAATGGGTGGCTGACCAACATTGCTCTTTTGT	654

```

QY 211 GlyAspProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCys 230
Db 655 GGTGATCAACTAAGGTACTATCTTTGGTGAATCTCTGCTGTTCTATGTCGTCACTGT 714
QY 231 HisLeuIleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGly 250
Db 715 CACATTTTGGACGAGCGGTGACACACTTACAAGGTAAGCCATTGTTTCAGAGCTGGT 774
QY 251 IleMetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGlu 270
Db 775 ATCATGCAATCTGGTGTATGTTTCCATCTGACGCCGTGACGGTATCTACGGTAACGAA 834
QY 271 IleTyrAspLeuPheValSerAlaGlyCysGlySerAlaSerAlaSerAspLysLeuAlaCys 290
Db 835 ATTTTTACATCTGTTGGCTTCCAAAGCTGTTGTTGTTCTGCTCTGCAAGTTGCTTGT 894
QY 291 LeuArgSerAlaSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeu 310
Db 895 TTGAGAGGTGTTTCTCTGACACTTTGGAAGACGCCACCAACACACCCCTGGTTCTTG 954
QY 311 AlaTyrSerSerLeuArgLeuSerTyrLeuProArgProArgProArgGlyLysAsnIleThrAsp 330
Db 955 GCTTACTCCTCTTAAGATTCTTACTTGTCCAGACGACGAGCGGTAAACATCACCGAC 1014
QY 331 AspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleIleGlyAsp 350
Db 1015 GACATGTACGCTTTGGTTAGAGAGGTAAAGTATGCCAACATCCCTGTATCATCGGTGAC 1074
QY 351 GlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnAlaGln 370
Db 1075 CAAACGACGAAGGTACCTTCTTGGTACTTCTTCTTCTTGAACGTTTACCACGTGATGCCAA 1134
QY 371 AlaArgAlaTyrPheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeu 390
Db 1135 GCCAGAGAATATTTCAAGCAATCTTTTGTCCACGCTACGACGCTGGAATCGACACTTGT 1194
QY 391 MetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsn 410
Db 1195 ATGACTGCTTACCCAGGTGACATCACTCAAGTTCTCCATTGACACTGGAACTTAAC 1254
QY 411 AlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHis 430
Db 1255 GCCTTGACCCACCAATTTCAAGAAATCTCTGCTGTTTGGGTGACTTGGGTTTACTTTG 1314
QY 431 AlaArgArgTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSerLys 450
Db 1315 GCTCGTAGATACCTTGAACCACTACACCGGTGTACCAAGTACTCTTCTTGTCTAAG 1374
QY 451 GlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAsp 470
Db 1375 CAATTGCTGTGTTGCCAGTTTGGGTACTTCCACTCCACGATATGCTTCCAAAGAC 1434
QY 471 TyrLeuLeuGlySerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAsp 490
Db 1435 TACTTGTGGTCTCTGTTCTTGTATCTACAACACGCTTTCATTGCTTTTGCCACTGAC 1494
QY 491 LeuAspProAsnThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerGln 510
Db 1495 TTGGACCCAAACACCGCGGTGTTGTTGTTAAGTGGCCAGAAATACACCTCTTCTTCTCAA 1554
QY 511 SerGlyAsnAsnLeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPhe 530
Db 1555 TCTGTGTAACAATCTGATGATGATCAACGCTTTGGGTGTTGATACCCGGTAAGGACAACTTC 1614
QY 531 ArgThrAlaGlyTyrAspAlaLeuMetThrAsnProSerSerPhePheVal 547
Db 1615 AGAACCGCGGTACGACGCTTGTCTCCAAACCCACCATCTTCTTCTTGT 1665

```

RESULT 5
CCLIPASE
LOCUS
DEFINITION

C. cylindracea mRNA for lipase I, partial cds. 1657 bp mRNA linear PLN 12-DEC-1997

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

X16712

X16712.1 GI:2548
lipase; lipase I.
Candida cylindracea
Candida cylindracea

REFERENCE
AUTHORS
TITLE

1 (bases 1 to 1657)
Kawaguchi, Y., Honda, H., Taniguchi-Morimura, J. and Iwasaki, S.
The codon CUG is read as serine in an asporogenic yeast *Candida*

cylindracea
Nature 341 (6238), 164-166 (1989)

JOURNAL

MEDLINE

PUBMED

FEATURES

source

Location/Qualifiers

1..1657

/organism="Candida cylindracea"

/mol_type="mRNA"

/strain="MS-5"

/db_xref="taxon:44322"

/clone="lambda CL115"

<1..1657

/note="for lipase 1"

<1..1617

/note="protein sequence is in conflict with the conceptual translation"

/codon_start=1

/codons=(seq:"ctg",aa:Ser)

/product="lipase 1"

/protein_id="CAA34684.1"

/db_xref="GI:2683215"

/db_xref="GOA:P20261"

/db_xref="SWISS-PROT:P20261"

/translation="SVAAAPTATLANGDTITGLNAINAEFLGIPFAEPVGNLRFKD
PVPVSGSLDGKFTSVGPGSQONPECTYEENLKAALDLVMQSKVFEAVSPSEDCI
TINVVRPPTGKAGANLPMWLWFGGFEVGTSTFPAQMITKSIAMGKPIIHVSVNY
RVSWGFLAGDEIKAESANGLKQRLGMQWADNIAAFGGDPTKVTIFGESAGMS
VMCHILNWDGNTYKGLPFRAGIMQSGAMVPSDAVDGIYNEIFDLIASNAGGSSAS
DKLACLRGSSDTLEDATNNTPGFLAYSLRSLYLPRPDGNIITDDMYALVREGKIAN
IPVTIGQNDGTEPFGTSSLNVTDAQAREYFKQSFVHSDAEITDILMTAYPDITOG
SPDTGILNALTPOFKRISAVIGDLGFTLARRYFLNHTVGTGTSFSLKQLSGPLVLG
TFHSDIVFDQYLLGSGSLIYNNAFIAFATDLPNTAGLLVKWPEYTTSSSQSGNNLNM
INAGLTYTGKDNFTAGTDALFSNPPSFV"

sig_peptide

mat_peptide

<1..12

13..1614

/product="lipase 1"

ORIGIN

Alignment Scores:

Pred. No.: 2,28e-193 Length: 1657

Score: 2416.00 Matches: 454

Percent Similarity: 90.50% Conservative: 32

Best Local Similarity: 84.54% Mismatches: 51

Query Match: 84.36% Indels: 0

DB: 8 Gaps: 0

US-09-943-857-4 (1-547) x CCLIPASE (1-1657)

QY 11 LeuGlySerValProThrAlaLysAlaLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAla 30

Db 4 GTGGTGTCTGCCCCCACCACCGCTCCGCCAACGGACACCATCACCGGTTCTCAACGCC 63

QY 31 IleIleAsnGluAlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArg 50

Db 64 ATCATCAACGAGCGCTTCTCGGCATTCCCTTTCGCGAGCGCGCGTCCGCAACCTCCGC 123

QY 51 PheLysAspProValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGly 70

Db 124 TTCAAGGACCCCGTCCGCTACTCCGCTCGCTCGATGCGGAGAGTTACGCTGTACGCG 183

QY 71 ProSerCysMetGlnGlnAsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAla 90

Db 184 CCGCTGTGATGCAGCAGACAGACCCCGGAGGACCTACGAGGAGAACCTCCCAAGGACGG 243

QY 91 LeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCys 110
 Db 244 CTCGACTTGGTGTGATGTCAGTCCAAGGTCTTTGAGGCGGTGCTCGCGCTGAGGAGACTGT 303
 QY 111 LeuThrIleAsnValValArgProProGlyThrLysAlaGlyAlaAsnLeuProValMet 130
 Db 304 CTCACCAATCAACGTGGTGGCGCGCGCCGAGCACCAGGCGGGTGCACACTCCCGGTGATG 363
 QY 131 LeuTriPheGlyGlyPheGluIleGlySerProThrIlePheProAlaGln 150
 Db 364 CTCTGGATCTTTGGCGCGGGTTTGAGGTGGGTGGCCACGACCTTCCCTCCCGCCAG 423
 QY 151 MetValThrLysSerValLeuMetGlyLysHileIleHisValAlaValAsnTyrArg 170
 Db 424 ATGATCACCAAGAGCATTCGTCGGCAAGCCCATCATCAGCGTGAGCGTCAACTACCGC 483
 QY 171 ValAlaSerTrpGlyPheLeuAlaGlyAspIleLysAlaGluGlySerGlyAsnAla 190
 Db 484 GTGTCTGCTGGGGGGTCTTTGGCTGGCGAGATCAAGGCCGAGGCGCAGTGCACAGCC 543
 QY 191 GlyLeuLysAspGlnArgLeuGlyMetGlnTrpValAlaAlaAspAsnIleAlaGlyPheGly 210
 Db 544 GGTTTGAAGGACCAAGCGCTTGGCATGTCAGTGGTGGCGGACCAACTTGGCGGTTTGGC 603
 QY 211 GlyAspProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCys 230
 Db 604 GGCAGCCCGACCAAGGTGACCATCTTGGCGAGCTGGCGGCGAGCATGTCGGTCACTGTC 663
 QY 231 HisLeuIleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGly 250
 Db 664 CAAITCTCTGGAAACGACGCGCAACAACGATCAAGGGCAAGCGCTCTTCGCGGGGC 723
 QY 251 IleMetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGlu 270
 Db 724 ATCATGAGCTGGGGCCATGGTGGCTGGACCGCTGGACCGCATCTACGGCAACGAG 783
 QY 271 IleTyrAspLeuPheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCys 290
 Db 784 ATCTTTGACTCTTGGCGTTCGAACGCGGGTGGCGGCGAGCGCCAGCAGCTTGGCTGC 843
 QY 291 LeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeu 310
 Db 844 TTGCGCGGTGTGTGAGCGACACGTTGGAGGACGCCCAACCAACACCCCTGGGTCTTG 903
 QY 311 AlaTyrSerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThrAsp 330
 Db 904 GCGTACTCTCTGTTGGTGTGTGTAACCTCCCGCGCGCGCGCGGTGAACATCACCGAC 963
 QY 331 AspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleIleGlyAsp 350
 Db 964 GACATGTAGCGCTTGGTGGCGGAGGCAAGTATGCCAACATCCCTGTGATCATCGCGGAC 1023
 QY 351 GlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnAlaGln 370
 Db 1024 CAGAACGACGAGGGACCTCTTTGGACCCCTGCTGTGAACGTGACCGAGTGCCTCCAG 1083
 QY 371 AlaArgAlaTyrPheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeu 390
 Db 1084 GCCCGGAGTACTTCAAGCAGCTGTGTTCACCGCCAGCGAGCGCGGAGATCGACGTTG 1143
 QY 391 MetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsn 410
 Db 1144 ATGACGGCGTACCCCNNGACATCACCCAGGGCGCTCGCGTTCGACACGGGTATCTCAAC 1203
 QY 411 AlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHis 430
 Db 1204 GGCCTCACCCCGAGTTCAAGAAATCTCGGGGTGCTCGGGGACCTTGGCTTTACGCTT 1263
 QY 431 AlaArgTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSerLys 450
 Db 1264 GTCGTGCTACTTCTTCAACCACTACACCGCGCGGCCAACAGTACTCTTCTCTCTGAAG 1323

QY 451 GlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAsp 470
 Db 1324 CAGCTCTCGGNTTGGCGGTGCTCGGAACTGTCACATCCCAACGACATGTTCCAGGAC 1383
 QY 471 TyrLeuLeuGlySerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAsp 490
 Db 1384 TACTTGTGGCGACGCGCTCGCTCATCTACAACAACGCGTTCATTGGCTTGGCACGGAC 1443
 QY 491 LeuAspProAsnThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGln 510
 Db 1444 TTGACCCCAACACCGCGGGTTGTGTGTAAGTGGCCGAGTACACACGAGCGCTCGAC 1503
 QY 511 SerGlyAsnAsnLeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPhe 530
 Db 1504 CTGGGCAACAACTGTATGATGATCAACGCTTGGCTTGTACACCGGCAAGGACAACTTC 1563
 QY 531 ArgThrAlaGlyTyrAspAlaLeuMetThrAsnProSerSerPhePheVal 547
 Db 1564 CGCACCGCGGCTACGACGCTGTCTTCTCCAAACCGCGCTGTCTTTGTG 1614
 RESULT 6
 A48377
 LOCUS A48377 Sequence 6 from Patent WO9603511. linear PAT 07-MAR-1997
 DEFINITION A48377
 ACCESSION A48377
 VERSION A48377.1 GI:2302168
 KEYWORDS
 SOURCE Candida rugosa
 ORGANISM Candida rugosa
 Bukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; mitosporic Saccharomycetales; Candida.
 REFERENCE 1 (bases 1 to 1733)
 AUTHORS Alibert, G., Mouloungui, Z. and Boudet, A.
 TITLE METHOD FOR PRODUCING FATTY ACIDS OR DERIVATIVES THEREOF FROM OIL
 JOURNAL PLANT
 PATENT: WO 9603511-A 6 08-FEB-1996;
 TOULOUSE INST NAT POLYTECH (FR)
 COMMENT Other publication AU 2984995 960222
 Other publication FR 2722798 960126.
 FEATURES
 source
 1. .1733
 /organism="Candida rugosa"
 /mol_type="unassigned DNA"
 /db_xref="taxon:5481"
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,94e-193 Length: 1733
 Score: 2415.00 Matches: 454
 Percent Similarity: 90.50% Conservative: 32
 Best Local Similarity: 84.54% Mismatches: 51
 Query Match: 84.32% Indels: 0
 DB: Gaps: 0
 US-09-943-857-4 (1-547) x A48377 (1-1733)
 QY 11 LeuGlySerValProThrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAla 30
 Db 37 GTGGTGTCTGCCCCACCGCCACCGCTCGCCACGCGGACCATCACCGGTCTCAACGCC 96
 QY 31 IleIleAsnGluAlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArg 50
 Db 97 ATCATCAACGAGCGCTTCTTCGGCATTCCTTTGCGAGCGCGCGGTGGGCAACTCTCGC 156
 QY 51 PheLysAspProValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGly 70
 Db 157 TTCAAGAGCCCCGTGCGGTACTCGCGCTCGCTCGATGCCAGAGTTCACGCTGTACGCGC 216
 QY 71 ProSerCysMetGlnGlnAsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAla 90
 Db 217 CGCTGTGTCATGCACAGAACCCCGAGGGCACCTACGAGGAGAACCTCCCCAAGCAGCG 276
 QY 91 LeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCys 110

```
Db 277 CTCAGCTTGGTGTGAGTCCAGTCCAGGTGTTTGGAGCGGTGCTGCGCTGAGGAGACTGT 336
Qy 111 LeuThrIleAsnValValArgProProGlyThrIysAlaGlyAlaAsnLeuProValMet 130
Db 337 CTCACCATCAAGTGGTGGCGCGCCGCGGACCAAGCGGGGTGCCAACCTCCCGGTGATG 396
Qy 131 LeuTrpIlePheGlyGlyPheGluIleGlySerProThrIlePheProAlaGln 150
Db 397 CTCGTGATCTTGGCGCGGTGTTGGAGTGGGTGGCCACGACCTTCCCTCCCGCCAG 456
Qy 151 MetValThrIysSerValLeuMetGlyIysHisIleIleHisValAlaValAsnTyrArg 170
Db 457 ATGATCACCAGACATGTCATGGCCAGCCCATCATCCACGAGGAGGTCACTACCGC 516
Qy 171 ValAlaSerTrpGlyPheLeuAlaGlyAspAspIleIysAlaGluGlySerGlyAsnAla 190
Db 517 GTGCGTCTGGGGGTCTTGGCTGGCGACGAGATCAAGGCGGAGGCGAGTGCACAGGCC 576
Qy 191 GlyLeuIysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGly 210
Db 577 GGTGTGAAGGACCGAGCTTGGCGATGCGATGAGTGGGTGGCGGACACATGTCGGCGTTGGC 636
Qy 211 GlyAspProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCys 230
Db 637 GGCACCGCACCAAGTGAACCATCTTTGGCGAGCTGGCGGCGAGCATGTGCTCATGTGC 696
Qy 231 HisLeuIleTrpAsnAspGlyAspAsnThrTrpIysGlyLysProLeuPheArgAlaGly 250
Db 697 CACATCTCTGAAACGACGCGGACAAACAGTACAGGCAAGCGCTCTTCGCGCGGGC 756
Qy 251 IleMetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGlu 270
Db 757 ATCATGACGTGGGGCCATGTGCGCTGGACGCGGTGGACGGCATCTACGGCACGAG 816
Qy 271 IleTyrAspLeuPheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCys 290
Db 817 ATCTTTGACCTTCTGGCGTGGACCGGGGCTGGCGGAGCGGCGGACGACAGCTTCGCTGC 876
Qy 291 LeuArgSerAlaSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeu 310
Db 877 TTGCGCGGTGTCTGAGCGACAGTGTGGAGGACGCCACCAACACCCCTGGGTCTTG 936
Qy 311 AlaTyrSerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThrAsp 330
Db 937 GGTACTCTCTGCTGGGTGTCTGACTCTCCCGCGCGGCGGCGGTGAACATCACCGAC 996
Qy 331 AspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleIleGlyAsp 350
Db 997 GACATGTACGCTTGGTGGCGGAGGCAAGTATGCCAACATCCCTGTGATCATCGCGCAC 1056
Qy 351 GlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnAlaGln 370
Db 1057 CAGAACGACGAGCGGACCTTCTTGGCACCCCTGCTGTGAACGTGACACGAGTATCTCAAC 1116
Qy 371 AlaArgAlaTyrPheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeu 390
Db 1117 GCCCGGAGTACTTCAAGCAGCTGTGTGTCACGCGGAGCGAGGATCGACAGTTG 1176
Qy 391 MetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsn 410
Db 1177 ATGACGGCGTACCCCGGCGACATCACCGAGGCGCTGCTCGGCGACCTTGGCTTACGCTT 1236
Qy 411 AlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHis 430
Db 1237 GCCCTCACCCCGCAGTCAAGAGATCCCTGGCGGTGCTCGGCGACCTTGGCTTACGCTT 1296
Qy 431 AlaArgTyrPheLeuAsnHisPheGlnGlyThrIysTyrSerPheLeuSerLys 450
Db 1297 GCTCGTGGTACTTCTCTCAACCACTACACCGGCGGACCAAGTACTCATCTCCCTCGTAAG 1356
Qy 451 GlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAsp 470
```

```
1357 CAGCTCCTGGCGTTCGCGGTGCTCGGAACGTTCCACTCCACGACGATCTGCTTCAGGAC 1416
Qy 471 TyrLeuLeuGlySerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAsp 490
Db 1417 TACTTGTGGGAGCGGCTCGCTCATCTACAAACGCGTTTCATTGGTTTGCCACGAC 1476
Qy 491 LeuAspProAsnThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGln 510
Db 1477 TTGACCCCAACACCGCGGGGTGTTGGTGAAGTGGCCGAGTACACGACGAGCTCGAG 1536
Qy 511 SerGlyAsnAsnLeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPhe 530
Db 1537 CTGGGCAACAACTTGATGATGATCAACGCTTGGGCTTGATACACCGCAAGGACACTTC 1596
Qy 531 ArgThrAlaGlyTyrAspAlaLeuMetThrAsnProSerSerPhePheVal 547
Db 1597 CGCACCGCGGTACGACGCGTGTCTCTCAACCGCGCTGTTCTTTGTG 1647

RESULT 7
CCLIP1 1733 bp DNA linear PLN 27-APR-1993
LOCUS C.cylindracea LIP1 gene for lipase.
ACCESSION X64703.1 GI:2544
VERSION X64703.1
KEYWORDS Lip1 gene.
SOURCE Candida cylindracea
ORGANISM Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 1733)
AUTHORS Longhi,S., Fusetti,F., Grandori,R., Lotti,M., Vanoni,M. and
Alberghina,L.
TITLE Cloning and nucleotide sequences of two lipase genes from Candida
cylindracea
JOURNAL Biochim. Biophys. Acta 1131 (2), 227-232 (1992)
MEDLINE 92305068
PUBMED 1610906
REFERENCE 2 (bases 1 to 1733)
AUTHORS Longhi,S.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-1992) S. Longhi, Universita' degli Studi di
Milano, Via Celoria 26, 20133 Milano, ITALY
COMMENT See also x64704
X66006, X66007 and X66008 are related sequences in the description
of C.cylindracea Lip1 and Lip2 sequences (x64703, x64704).

FEATURES
Location/Qualifiers
1..1733
/organism="Candida cylindracea"
/mol_type="genomic DNA"
/strain="ATCC 14830"
/db_xref="taxon:44322"
/clone="A"
1..1650
/gene="LIP1"
1..1650
/gene="LIP1"
/gene="LIP1"
/EC_number="3.1.1.3"
/codon_start=1
/product="triacylglycerol lipase"
/protein_id="CAA45957.1"
/db_xref="GI:1325987"
/db_xref="GOA:P20261"
/db_xref="SWISS-PROT:P20261"
/translation="MELALALLIAASAAAPTATLANGDTTGLNAINAEAPLGPFA
EPVGNLRFKDPVPYSGSDGQKFTLYGLCMQNPGETYENLEPKALDLVMQSKVF
EAVLPLSDDCLTINVRPPTGKAGANLPMWIFGGGEVGGTSTFPQAQTKTSIAM
GKPLIHVSNNVSVSWGFLAGDEIKAEGSANAGLQORLQMWADVADNIAFGGDPKTV
TIFGELGASMSVMCHILNDGNTYKGLPEAGIMOLGAMVPLDAVDGIVGNEIPDL
IASNAGCGSADKLACLKGLVSDTLEDTATNNTPGFLAYSSLRLLYLPRDGVNITDM
VALVRKYANTPVIFGQNDDEGTFGLLNLTVDTAQAREYFKQLFVHASDAEIDTL
MTAPYGDITQGLPFDNTIQKRIILNALTQKRIALVGLDGLTHARRYFLNHVTGGTKYSL
LAKLLGLVLGTFHSDNIVFDYLLGSGSLIYNNAFIAPATDLPNTAGLLVKWPEYT
SSLQLGNLMLNALGLYTGKDNFTAGYDALFSNPPLFFV"
```

```

sig_peptide      1..45
                  /gene="LIP1"
mat_peptide      46..1647
                  /note="triacylglycerol lipase"
                  /gene="LIP1"
                  /product="triacylglycerol lipase"
                  /EC_number="3.1.1.3"

ORIGIN
Alignment Scores:
Pred. No.:      2,945-193      Length:      1733
Score:          2415.00      Matches:     454
Percent Similarity: 90.50%      Conservative: 32
Best Local Similarity: 84.54%      Mismatches:  51
Query Match:    84.32%      Indels:      0
DB:             8      Gaps:        0

US-09-943-857-4 (1-547) x CCLIP1 (1-1733)

QY 11 LeuGlySerValProThrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAla 30
   ::::
Db 37 GTGGTGTCTGCCCCACCGCCAGCTCGCCACACGGCGACACCATCACCGTCTCAACGCC 96

QY 31 IleIleAsnGluAlaPheLeuGlyIleProPheAlaGluProValGlyAsnLeuArg 50
   ::::
Db 97 ATCATCAACGAGCGGTCTCTCGGCATTCTCGGCATTCTCGGCAGCGCGGTGGGCAACCTCCGC 156

QY 51 PheLysAspProValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGly 70
   ::::
Db 157 TTCAGGACCCCGTCCCGTACTCCGGCTCGCTGATGGCCAGAAGTTACGCTGATACGGC 216

QY 71 ProSerCysMetGlnGlnAsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAla 90
   ::::
Db 217 CCGCTGTGATGATGAGTCCAGGTCCTGAGGCGGTCTCGCGTGGAGGAGCACTCCCGAGGCGAG 276

QY 91 LeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCys 110
   ::::
Db 277 CTCGACTTGGTATGATGAGTCCAGGTCCTGAGGCGGTCTCGCGTGGAGGAGCACTGT 336

QY 111 LeuThrIleAsnValValArgProProGlyThrLysAlaGlyAlaAsnLeuProValMet 130
   ::::
Db 337 CTCACCATCAACGTGTGCGCGCGCGCGGACCAAGCGGGTGCCAACTCCCGCGTGATG 396

QY 131 LeuTrpIlePheGlyGlyPheGluIleGlySerProThrIlePheProAlaGln 150
   ::::
Db 397 CTCGTGATCTTTGGCGCGGGTTTGAGTGGGTGGCCACCACTTCCTCCCGCCGAG 456

QY 151 MetValThrLysSerValLeuMetGlyLysHisIleIleHisValAlaValAsnTyrArg 170
   ::::
Db 457 ATGATCAACCAAGAGCATTTGCCATGGGCAGGCCCATCATCCAGTGCAGCGTCAACTACCGC 516

QY 171 ValAlaSerTrpGlyPheLeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsnAla 190
   ::::
Db 517 GTGTGTGTGGGGGTTCTTGCTGGCGACAGATCAAGCGCGAGGCGAGTGCACACGCC 576

QY 191 GlyLeuLysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGly 210
   ::::
Db 577 GGTTTGAAGGACCAAGCGCTTGGCATGCGAGTGGGTGGCGGACAACTTGGCGCGTTTGGC 636

QY 211 GlyAspProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCys 230
   ::::
Db 637 GCGCAGCCCGACCAAGGTGACCATCTTTGGCGAGCTGGCGGCGAGCATGTGCGGTCTATGTC 696

QY 231 HisLeuIleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGly 250
   ::::
Db 697 CACATTCTGGAACGACGCGGACACACGTAACAGGGCAACCGCTCTTCGCGCGGGC 756

QY 251 IleMetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGlu 270
   ::::
Db 757 ATCATGCAGTGGGGCCATGTGTCGCTGGACCGCGTGGACGGCATCTACGGCAACGAG 816

QY 271 IleTyrAspLeuPheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCys 290
   ::::

```

```

Db 817 ATCTTTGACCTCTTGGCGTCGAACCGGGCTGCGGAGCGGCCAGCAGCAGCTTGGCGTGC 876
QY 291 LeuAqSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeu 310
   ::::
Db 877 TTGGCGGGTGTCTGAGCGACAGCTTGGAGAGCGCCCAACCAACACCCCTTGGGTCTTG 936
QY 311 AlaTyrSerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThrAsp 330
   ::::
Db 937 GCGTACTCTCGTTGGGTGTGTGTACCTCCCGCGCGCGCGCGTGAACATCACCGAC 996
QY 331 AspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleLeuGlyAsp 350
   ::::
Db 997 GACATGTACGCTTGGTGGCGGAGGCAAGTATGCCAACATCCTCTGTATCATCGCGGAC 1056
QY 351 GlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnAlaGln 370
   ::::
Db 1057 CAGACGACGAGGCGACCTTCTTTGGCACCTCTGTGTGAACGTGACCGAGTATGCCAG 1116
QY 371 AlaArgAlaTyrPheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeu 390
   ::::
Db 1117 GCCCGGAGTACTTCAAGCAGCTGTGTTCACGCGCGAGCGCGGAGATCGACAGTTG 1176
QY 391 MetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsn 410
   ::::
Db 1177 ATGACGGCGTATCCCGCGGACATCACCCAGGCGCTGCGCTTCGACACGGGTATCTCAAC 1236
QY 411 AlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHis 430
   ::::
Db 1237 GCCCTCACCCCGAGTTCAAGAGAATCTTGGCGGGTGTCTCGCGGACCTTGGCTTTAGCCTT 1296
QY 431 AlaArgTyrTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSerLys 450
   ::::
Db 1297 GCTGTGTGCTACTTCTCAACCACTACACCGCGCGGACCAAGTACTCATCTCTCTGAAG 1356
QY 451 GlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAsp 470
   ::::
Db 1357 CAGCTCTGGGGTGTGCGGTGTCTCGAAGCTTCCACTCCAAACGACATCTCTTCCAGGAC 1416
QY 471 TyrLeuLeuGlySerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAsp 490
   ::::
Db 1417 TACTTGTGGGCGAGCGGTCTCGCTCATCTACAAACAGCGGTTCATTGCGTTTGCACGGAC 1476
QY 491 LeuAspProAsnThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGln 510
   ::::
Db 1477 TTGGACCCCAACACCGCGGGTGTGTGGTGAAGTGGCGCGGAGTACACAGCAGCCTGCAG 1536
QY 511 SerGlyAsnAsnLeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPhe 530
   ::::
Db 1537 CTGGGCAACAACTTGATGATGATCAACGCTTGGGCTTGTACACGGCGAAGGACAACCTTC 1596
QY 531 ArgThrAlaGlyTyrAspAlaLeuMetThrAsnProSerSerPhePheVal 547
   ::::
Db 1597 CGCACCGCGGTACAGCGGTGTCTTCCAAACCCCGCGCTGTCTTCTTGTG 1647

RESULT 8
AB1172
LOCUS      AB1172
DEFINITION Sequence 2 from Patent WO9914338.
ACCESSION AB1172
VERSION    AB1172.1
KEYWORDS   GI:6731631
SOURCE     unidentified
           unidentified
           unclassified.
REFERENCE 1 (bases 1 to 1605)
AUTHORS   Schmidt-Dannert,C. and Schmid,R.
TITLE     TOTAL SYNTHESIS AND FUNCTIONAL OVEREXPRESSION OF A CANDIDA RUGOSA
           LIP1 GENE CODING FOR A MAJOR INDUSTRIAL LIPASE
JOURNAL   Patent: WO 9914338-A 2 25-MAR-1999;
           SCHMIDT DANNERT CLAUDIA (DE); SCHMID ROLF (DE)
FEATURES
           Location/Qualifiers
           1..1605
           /organism="unidentified"

```

/mol_type="unassigned DNA"
/db_xref="taxon:32644"

ORIGIN

Alignment Scores:

Pred. No.: 1,87e-191 Length: 1605
Score: 2393.00 Matches: 458
Percent Similarity: 90.71% Conservative: 30
Best Local Similarity: 85.13% Mismatches: 35
Query Match: 83.55% Indels: 16
DB: 6 Gaps: 2

US-09-943-857-4 (1-547) x A81172 (1-1605)

```

Qy 11 LeuGlySerValProThrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAla 30
Db 37 GTTGCTGCTGCCCAACCGGCACCTTTGGCTAAACGGTGACACCATCAACGGTTGAACGCC 96
Qy 31 IleileAsnGluAlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArg 50
Db 97 ATCATCAACGAGCCTTCTGGGTATTCATTTCCATTTCCGCAACACACCGTTGGTAACCTGAGA 156
Qy 51 PheLysAspProValProThrSerGlySerLeuAsnGlyGlnLysPheThrSerThrGly 70
Db 157 TTCAGGACCCAGTCCATATCCGGTTCCTTGGATGGTCAAAAAGTTCACTTCTTACGGT 216
Qy 71 ProSerCysMetGlnGlnAsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAla 90
Db 217 CCATCTTGTATGCAACAAACCCGAGAGGTACCTACGAGAAACCTTGGCCAAAGGCAGCT 276
Qy 91 LeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCys 110
Db 277 TTAGATCTGGTTATGCATCCAACT-----CCATCTTCTGAAGACTGT 320
Qy 111 LeuThrIleAsnValValArgProProGlyThrLysAlaGlyAlaAsnLeuProValMet 130
Db 321 TTGACCATTAATGTGTAGACACCCCGGCAAAAGGCTGGTGCCAACTTGGCCAGTATG 380
Qy 131 LeuThrIlePheGlyGlyGlyPheGluIleGlySerProThrIlePheProAlaGln 150
Db 381 TTGGGATCTTTGGTGGGTTTGAAGTTGGTGGTACTAGTACCTTCCCTCCAGCCCAA 440
Qy 151 MetValThrLysSerValLeuMetGlyLysHisIleIleHisValAlaValAsnThrArg 170
Db 441 ATGATTACCAAGTCTATTGCTATGGTGAAGCCATCATCCAGCTTCTGTCACTACAGA 500
Qy 171 ValAlaSerTrpGlyPheLeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsnAla 190
Db 501 GTCTCGAGCTGGGTTTCTTGGCTGTCGCAAAATCAAGGCCGAAGGTTCTGCCAACGCC 560
Qy 191 GlyLeuLysAspGlnAtcLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGly 210
Db 561 GGTGTGAAGGACCAAGGATGGGTATGCAATGGGTGGCTGACAACTTGGCTTTGGT 620
Qy 211 GlyAspProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCys 230
Db 621 GGTGATCCAACTAAGTTACTATCTTTGGTGAATCTGCTGGTCTCTATGCTCGTCACTGT 680
Qy 231 HisLeuIleTrpAsnAspGlyAspAsnThrTrpLysGlyLysProLeuPheArgAlaGly 250
Db 681 CACATTTTGTGCAACGCGGTGACAACTTACAGGGTAAAGCCCATTTGTCAGAGCTGGT 740
Qy 251 IleMetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTrpGlyAsnGlu 270
Db 741 ATCATGCAATCTGGTGTATGGTTCCATCTGACGCGCTGACGGTATCTACGGTAACGAA 800
Qy 271 IleTrpAspLeuPheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCys 290
Db 801 ATTTTGTACTTGTGGCTTCCAAACGCTGGTTCTGGTCTCTGCTCTGACAAAGTTGGCTTGT 860
Qy 291 LeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeu 310
Db 861 TTGAGAGGTGTTTCTTCTGACACTTTGGAAGACGCCCAACAAACACCCCTGGTTCTTGT 920

```

```

Qy 311 AlaTrpSerSerLeuArgLeuSerTrpLeuProArgProAspGlyLysAsnIleThrAsp 330
Db 921 GCCTTACTCCTCTTAAGATTGTTCTTACTGCCAAGACACGACGGTGTAAACATCACCGAC 980
Qy 331 AspMetTrpLysLeuValArgAspGlyLysTrpAlaSerValProValIleIleGlyAsp 350
Db 981 GACATGTACGCTTTGGTTAGAGAAGGTAAAGTATGCCAACATCCCTGTATTCATCGGTGAC 1040
Qy 351 GlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnAlaGln 370
Db 1041 CAAACACGACGAGGTACCTTCTTGGTACTTCTTCTTGAACGTTACCATCTGATGCCAA 1100
Qy 371 AlaArgAlaTrpPheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeu 390
Db 1101 GCCAGAGAAATATTTCAAGCAATCTTTGTCCCGTAGCGACGCTGAATTCGACACTTGT 1160
Qy 391 MetAlaAlaTrpProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsn 410
Db 1161 ATGACTGCTTACCCAGGTGACATCAAGGTCTCTCATTTGACACTGGAATTCATAAC 1220
Qy 411 AlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHis 430
Db 1221 GCCTTGACCCCAATTCACAGAAATCTCTGCTGTTTGGTGACTTGGGTTTACTTTG 1280
Qy 431 AlaArgArgTrpPheLeuAsnHisPheGlnGlyThrLysTrpSerPheLeuSerLys 450
Db 1281 GCTCGTAGATACTTCTTGAACCACTACACCGGTGGTACCAAGTACTCTTCTTCTCTAAG 1340
Qy 451 GlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAsp 470
Db 1341 CAATTGCTGCTGGTTGGCCAGTTTGGGTACTTTCACCTCAACGATATCGTCTTCCAGAC 1400
Qy 471 TyrLeuLeuGlySerGlySerValIleTrpAsnAsnAlaPheIleAlaPheAlaThrAsp 490
Db 1401 TACTTGTGGTCTGGTCTCTGATGATCAACACGCTTTCATTTGCTTGGTCCACTGAC 1460
Qy 491 LeuAspProAsnThrAlaGlyLeuLeuValAsnTrpProLysTrpThrSerSerPheVal 510
Db 1461 TTGACCCAAACACCGCGGT-----TTCTTCTCTCTCA 1493
Qy 510 nSerGlyAsnAsnLeuMetMetIleAsnAlaLeuGlyLeuTrpThrGlyLysAspAsnPh 530
Db 1494 ATCTGGTAAACAACTTGATGATGATCAACGCTTGGGTTTGTACACCGGTAAAGCAACTT 1553
Qy 530 eAtqThrAlaGlyTrpAspAlaLeuMetThrAsnProSerSerPhePheVal 547
Db 1554 CAGAACCGCGGTACGACGCTTGTCTCCAACCCCATCTTCTTCTTGT 1605

RESULT 9
AX709921 1532 bp DNA linear PAT 10-APR-2003
LOCUS Sequence 3 from Patent EP1288294.
DEFINITION AX709921
ACCESSION AX709921.1 GI:29786302
VERSION AX709921.1
KEYWORDS Candida rugosa
SOURCE Candida rugosa
ORGANISM Candida rugosa
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE
1
AUTHORS Tang, S.J., Lee, G.C. and Shaw, J.F.
TITLE Recombinant candida rugosa lipases
JOURNAL Patent: BP 1288294-A 3 05-MAR-2003;
Academia Sinica (TW)
FEATURES
source Location/Qualifiers
1..1532
/organism="Candida rugosa"
/mol_type="unassigned DNA"
/db_xref="taxon:5481"

ORIGIN
Alignment Scores:

```


Pred. No.: 1,93e-191 Length: 1532
 Score: 2392.50 Matches: 502
 Percent Similarity: 94.37% Conservatives: 1
 Best Local Similarity: 94.18% Mismatches: 9
 Query Match: 83.54% Indels: 26
 DB: 6 Gaps: 16

US-09-943-857-4 (1-547) x AX709921 (1-1532)

QY 15 ProThAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAlaIleIleAsnGlu 34
 DB 2 CCCACGCCCAAGCTCGCCAAAGCGGACACCATCACCGGTCTCAACGCCCATCATCAACGAG 61
 QY 35 AlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArgPheLysAspPro 54
 DB 62 GCGTTCCTCGGCATTCCTTTGCGAGCGCGCGTGGCAACCTCCGCTTCAAGGACCT 121
 QY 55 ValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyProSerCysMet 74
 DB 122 GTGCCGTACTTGGCTCGCTCAACGCCCAAGATT-ACT---TACGCGCGCG---TGCAATG 174
 QY 75 GlnGlnAsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAlaLeuAspLeuVal 94
 DB 175 CAGCAGAACCCCGAGGCGCAGTTTGAGAGAACCTTGGCAAGACGGCAGCTCGACTTGGTG 234
 QY 95 MetGlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsn 114
 DB 235 ATGCAGTCCAAGGTGTCCAGGCGGTGCTTCCCGAGAGTGAGGACTGCCTCACCATCAAC 294
 QY 115 ValValArgProProGlyThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePhe 134
 DB 295 GTGGTGGCGCGCGCGGACCAAGGCGGCGGCGCAACCTCCCGGTGATGCTCTGGATCTTT 354
 QY 135 GlyGlyGlyPheGluIleGlySerProThrIlePheProProAlaGlnMetValThrLys 154
 DB 355 GCGGFTGGGTTTGAGATCGGACGCCCCACCATCTTCCCTCCCGCCAGATGGTCAACAAG 414
 QY 155 SerValLeuMetGlyLysHisIleIleHisValAlaValAsnTyrArgValAlaSerTrp 174
 DB 415 AGTGTCTCATGGCAAGCC-ATCATCCAGTGGCGGTCACTACCGTGTGTGCTCGTGG 473
 QY 175 GlyPheLeuAlaGlyAspAspIleLysAlaGlyGlySerGlyAsnAlaGlyLeuLysAsp 194
 DB 474 GGGTCTTGGTGTGATGATCAATCAAGCGCGAGCGGAGCGGCAACCGCGCTTGAAGGAC 533
 QY 195 GlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSer 214
 DB 534 CAGCGTTTGGGCATGAGTGGTGGGAGACAAATTTGCCGGTTCGGCGGAGCCCGAGC 593
 QY 215 LysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuGlyHisLeuIleTrp 234
 DB 594 AAGGTGACATC-TTTGGCGAG--GCGGCGAGCATGTCCGTGTGTGCGCACCTCATCTGG 649
 QY 235 AsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSer 254
 DB 650 AACGACGGCGCAACACGTACAGGCGCAAGCGCTTGTCCGCGCGGCATCATGCGAG-- 706
 QY 255 GlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGluIleTyrAspLeu 274
 DB 707 GGAGCCATGTGCGG--GACCGGTGGAGCGGACGATGCGGCAACGAGATCTACGACCTC 763
 QY 275 PheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCysLeuArgSerAla 294
 DB 764 TTTGTCTCGAGTGTGGTGTGGCAGCGCCAGCAGCATCGCGTGTGCGCAGTGG 823
 QY 295 SerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeuAlaTyrSerSer 314
 DB 824 ---AGCGACACCTTGTCTGATGTCACCAACACACTCTCTGGTCTTGTGGTACTCTCTCG 880
 QY 315 LeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThrAspAspMetTyrLys 334
 DB 881 TTGCGGTG-----TACTCCGCGCGCGGAGCAACATCACCATGATGATGATCAAG 934

QY 335 LeuValArgAspGlyLysTyrAlaSerValProValIleIleGlyAspGlnAsnAspGlu 354
 DB 935 TTGGTGGCGACGCGCAAGTATGCAAGGTTCCCGTGATCATTTGGCGACCAAGACGAG 994
 QY 355 GlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnAlaGlnAlaArgAlaTyr 374
 DB 995 GGCACCATCTTTGGCTC-----TTGAACGTGACCAAGTAATGTCAGGCCGCTGTAC 1047
 QY 375 PheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaIleTyr 394
 DB 1048 TTCAAGCAG---TTCATCCACGCCGAGCGGAGATCGACACCTTGATGGCGGTAC 1104
 QY 395 ProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsnAlaLeuThrPro 414
 DB 1105 CCCCAGACATCACCCAGGCT---CGTTTCGACAC---GGT-----TCAACGCTCACCCG 1154
 QY 415 GlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHisAlaArgArgTyr 434
 DB 1155 CAGTTCAGAGAGATC---GCGGTGCTCGGCGACCTTGATTCATCCACGCCCGCGCTAC 1211
 QY 435 PheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSerLysGlnLeuSerGly 454
 DB 1212 TTCTCAACACACTTCCAGGCGCGCAAGTACTCGTTCTC---AAGCAGCTC---GGG 1265
 QY 455 LeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAspTyrLeuLeuGly 474
 DB 1266 TTGCCAATCATGGGACCTTCCATGCCACGACATTTGTGGCAGGACTACTTGTGGGA 1325
 QY 475 SerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsn 494
 DB 1326 AGCGGAGCGGTCTATCTACCAACACGCGTTTATCGCTTCCGCTCGCCACCGACTTGGACCCCAAC 1385
 QY 495 ThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerGlnSerGlyAsnAsn 514
 DB 1386 ACCGCGGGGTGTGTGTGTAAGTGGCCCAAGTACACCGAGC-----CAGGCGCAAC 1439
 QY 515 LeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGly 534
 DB 1440 TTGATGATCATCAACGCTTGGCTTGTACCGGCAAGGACAACTTCGACCGCTGGC 1499
 QY 535 TyrAspAlaLeuMetThrAsnProSerSerPhePheVal 547
 DB 1500 TAGCAGCGTGTATGATCAACACCG-----TCTTTGTG 1532

RESULT 10

LOCUS CCLIP5 1812 bp DNA linear PLN 30-JUN-1993
 DEFINITION C.cylindracea LIP5 gene.
 ACCESSION X66008.S55942
 VERSION X66008.1 GI:296937
 KEYWORDS lip5 gene; lipase.
 SOURCE Candida cylindracea
 ORGANISM Candida cylindracea
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; mitosporic Saccharomycetales; Candida.
 REFERENCE 1 (bases 1 to 1812)
 AUTHORS Lotti,M., Grandori,R., Fusetti,F., Longhi,S., Brocca,S.,
 Tramontero,A. and Alberghina,L.
 TITLE Cloning and analysis of Candida cylindracea lipase sequences
 JOURNAL Gene 124 (1), 45-55 (1993)
 MEDLINE 93178975
 PUBMED 840480
 REFERENCE 2 (bases 1 to 1812)
 AUTHORS Longhi,S.
 TITLE Direct Submission
 JOURNAL Submitted (06-MAY-1992) S. Longhi, Universita' degli Studi di
 Milano, Via Celoria 26, 20133 Milano, ITALY
 COMMENT See X66006-8, X16712, X64703 & X64704
 X66006, X66007 and X66008 are related sequences in the description
 of C.cylindracea Lipi and Lip2 sequences (x64703, x64704).
 FEATURES
 Location/Qualifiers
 1..1812
 /organism="Candida cylindracea"


```

/molecule="genomic DNA"
/strain="ATCC 14830"
/db_xref="taxon:4322"
/clone_lib="SACI genomic"
116..1765
/gene="LIP5"
116..1765
/gene="LIP5"
/EC_number="3.1.1.3"
/codon_start=1
/product="triacylglycerol lipase"
/protein_id="CAA46807.1"
/db_xref="GI:1325991"
/db_xref="GOA:P32949"
/db_xref="SWISS-PROT:P32949"
/translation="MKLALALLIASVAAPATATLANGDTTIGLNALINEAFGLIPFA
EPPVGNLRFDPVPRGSLNGSFTAYGPLCMQNPREGTYEENLPKVALDLVMQSKVP
QAVLPNSBCLITINVPPGKAGANLPVLMWIFGGFPEIGSPTTFPPAOMVSKSLM
GKPTIHVAVNYRLASFGFLAGPDIKAGSSNAGLKQORLGMQWADVADNAGFGGDFSKV
TIFELAGSSNVLCHLLWNGDNTYKQKPLFRAGIMOLGAMVPLDPVDVGTGTQIYDT
LVALTGSSASNKLAICLRLGLTQALLDATNDTPGFLSVTSRLLYLPRPDGANTIDDM
YKVRDGYASVPVIGDNDGELFGLLNLTTFEADAEYLRKLFHATDADITAL
KAAVPSDVDTQGLPDTFNLALTQPKRINAVLGLTFTLSRRYFLNHYTGPKYSFL
LKQLGLPILGTGHANDIVQHFLLGSGSVYNNAFATADLDPNTAGLLVQWFKST
SSLQAGDLMQISALGLYTGKONFRTAGYNALFADPLHFFV"
116..250
/sig_peptide
251..1762
/mat_peptide
/product="triacylglycerol lipase"
/EC_number="3.1.1.3"

ORIGIN
Alignment Scores:
Pred. No.: 1.61e-187 Length: 1812
Score: 2347.00 Matches: 445
Percent Similarity: 88.83% Conservative: 32
Best Local Similarity: 82.87% Mismatches: 60
Query Match: 81.95% Indels: 0
DB: 8 Gaps: 0

US-09-943-857-4 (1-547) x CCLIP5 (1-1812)
Qy 11 LeuGlySerValProThrAlaLysLeuAlaAenGlyAspThrIleThrGlyLeuAsnAla 30
Db : : : : :
Db 152 GTGGTGTGTCGCCCCCACCAGCGTTCGCCAACAGCGGACACCATCACCAGTCTCAACGCC 211
Qy 31 IleIleAsnGluAlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArg 50
Db : : : : :
Db 212 ATCATCAAGAGCGGTCTCTCGCATTCCTTTGCGAGCGCGCGTGGGCAACCTCCGC 271
Qy 51 PheLysAspProValProTyrSerGlySerLeuAenGlyGlnLysPheThrSerTyrGly 70
Db : : : : :
Db 272 TTCAAGAGACCTGTGCGGTACCGGTCTCTCAACGGTCAATCTTCACCGCGTACCGGT 331
Qy 71 ProSerCysMetGlnGlnAsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAla 90
Db : : : : :
Db 332 CCGCTGTGATGACGAGAACCCGAGGGCACCTACGAGAGAACCTCCCCAAGGTGGCG 391
Qy 91 LeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCys 110
Db : : : : :
Db 392 CTTGACTTGTGATGCAGTCCAAAGGTGTTCCAGGTGTTCTCCCCAACAGCGAGACTGC 451
Qy 111 LeuThrIleAsnValValArgProProGlyThrLysAlaGlyAlaAsnLeuProValMet 130
Db : : : : :
Db 452 CTCACCATCAACGTGTGCGCGCGCGGACCAAGGGGGGCGCAACCTCCCGGTGATG 511
Qy 131 LeuTyrIlePheGlyGlyGlyPheGluIleGlySerProThrIlePheProAlaGln 150
Db : : : : :
Db 512 CTTGTGATCTTTGGCGGTGGGTGTGAGATCGGACGCCAGCCACCATCTTCCCTCCCGTCA 571
Qy 151 MetValThrLysSerValLeuMetGlyLysIleIleHisValAlaValAsnTyrArg 170
Db : : : : :

```

```

Db 572 ATGGTCTCCAAAGAGTGTGCTCATGGGCAAGCCCAATCATCCAGTGGCGCGTCAATCCGC 631
Qy 171 ValAlaSerTrpGlyPheLeuAlaGlyAspIleLysAlaGluGlySerGlyAsnAla 190
Db : : : : :
Db 632 TTGGCGTCTTTGGTTTCTTGGCCGTCGGGACATCAAGGCGGAGGAGTCCCAATGCC 691
Qy 191 GlyLeuLysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGly 210
Db : : : : :
Db 692 GGCCTCAAGGACCAAGCGCTTGGGCATGCAGTGGGTGGCAGACAACATTCGCGGTTCGCG 751
Qy 211 GlyAspProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCys 230
Db : : : : :
Db 752 GGCACCCGAGCAAGGTGACCATCTTTGGTGAGTGGCGGCGGACATGTCGTTGTTGTC 811
Qy 231 HisLeuIleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGly 250
Db : : : : :
Db 812 CACCTTCTCTGGAATGGCGGCGACACACATGTAAGGCGCAACCGTGTTCGCGCGGCG 871
Qy 251 IleMetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGlu 270
Db : : : : :
Db 872 ATCATGCGAGTGGGAGCATGTCGCGCTGGACCCGCTGGACGCGCACCTATGGAAACCCAA 931
Qy 271 IleTyrAspLeuPheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCys 290
Db : : : : :
Db 932 ATCTATGACACGTTGGTGGCTCTGACGCGGCTGCAGAGTGCAGCAACAGCTTCGCTGC 991
Qy 291 LeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeu 310
Db : : : : :
Db 992 TTGGTGTGCTTCTGACTCAGGCATGCTCGATGCCCAACAGCACCCCTGGGTCTTG 1051
Qy 311 AlaTyrSerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThrAsp 330
Db : : : : :
Db 1052 TCGTACACCTGTTGCGGTTCGTGTACCTCCCGCGCGCGGCGGCGCAACATCACCGAT 1111
Qy 331 AspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleIleGlyAsp 350
Db : : : : :
Db 1112 GACATGTACAAGTGGTACGCGAGCGCAAGTATGCAACGCTTCCGCGTATCATTCGCGCAC 1171
Qy 351 GlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnAlaGln 370
Db : : : : :
Db 1172 CAGAACGACGAGGGCTTCTGTTGGTCTCTCGTGTGTAACACACCCAGGAGCGCGAC 1231
Qy 371 AlaArgAlaTyrPheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeu 390
Db : : : : :
Db 1232 GCGAGGACATACCTCAGAAAGCTGTTCATCCAGCCACCGAGCGCGGATATCACCGCATG 1291
Qy 391 MetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsn 410
Db : : : : :
Db 1292 AAGCGCGGTACCCGAGCATGTACCCAGGGTCTGCGGTTCGACGAGGCGATTCACAC 1351
Qy 411 AlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHis 430
Db : : : : :
Db 1352 GCCCTTACACCCAGCTCAAGCGGATCAATGCTGTGCTGGCGACCTCACCTTACCCTC 1411
Qy 431 AlaArgTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSerLys 450
Db : : : : :
Db 1412 TCGCGCGGTACTCTCTCAACCACTACACCGGTGTCCTCAAGTACTCGTTCCTCTGAAG 1471
Qy 451 GlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAsp 470
Db : : : : :
Db 1472 CAGCTTCTGGGCTTCCCATCTTCGCGACGTTCCAGCGGACGACATGTTGTGGCAGCAC 1531
Qy 471 TyrLeuLeuGlySerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAsp 490
Db : : : : :
Db 1532 TTTTGTGGGCGAGCGGCGGCGGTCATCTACAACACGCGTTCATCGCTTTCGCCACGAC 1591
Qy 491 LeuAspProAsnThrAlaGlyLeuValAsnTrpProLysTyrThrSerSerSerGln 510
Db : : : : :
Db 1592 TTGGACCCCAACACCGCGGCTTGTGTGGTGCAGTGGCGGCAAGTCCACGAGCGCTGCAG 1651
Qy 511 SerGlyAsnAsnLeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPhe 530
Db : : : : :
Db 1652 GCGGGGACCAACTTATGACAGATCATGCTTGGGCTGTATACCCGCGGCAAGGCAACTTC 1711

```



```
Db 1155 CAGAACGACGAGGACAGGTTTGGCTGCTGTGTGAACGTACTACGATGCTCAG 1214
Qy 371 AlaArgAlaTyPheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThrIleu 390
Db 1215 GCACGCCAGTACTTCAAGGAAGTCTTCATCCACGCCACGACGCGGAGATCGACACCTTG 1274
Qy 391 MetAlaAlaTyPProGlnAspIleThrGlnGlySerProPheAspThrGlyValIleuAsn 410
Db 1275 ATGCGCGGTACCCACGACATCACCCAGGGTCTGCCGTTTCGACACCGGATCTTCAAC 1334
Qy 411 AlaLeuThrProGlnPheLysArgIleSerAlaValIleuGlyAspLeuAlaPheIleHis 430
Db 1335 GCCATCACCCGCGATTCACACGAGATTGACGGGTGCTTGGTACCTTGGTCACTCTC 1394
Qy 431 AlaArgArgTyPheLysHisPheGlnGlyThrLysTyPheLysSerPheIleuSerLys 450
Db 1395 CCCCGCGCTACTTCCCTCAACCACTTCCAGGCGCGCACCAAGTACTCGTCTCCTGAAG 1454
Qy 451 GlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTriPAlaAsp 470
Db 1455 CAGCTTCTGGGTTCGCCGTGATTGGCACCCACCCACGACGACGATTTGTGGCAGAC 1514
Qy 471 TyrLeuLeuGlySerGlySerValIleTyPheHisAlaAsnAlaPheIleAlaPheAlaThrAsp 490
Db 1515 TTTTGGTGAGCCACGACGCGCGTGTACACCAACGCGTTTATTGCTTTGCCAACGAC 1574
Qy 491 LeuAspProAsnThrAlaGlyLeuLeuValAsnTrpProLysTyPheLysSerSerGln 510
Db 1575 CTCGACCCGAAACAGGCGGTTTGTGTGAACCTGGCCCAAGTACACCGACGCGTGCAG 1634
Qy 511 SerGlyAsnAsnLeuMetMetIleAsnAlaLeuGlyLeuTyPheLysAspAsnPhe 530
Db 1635 CTGGGCAACAACTTGTGAGATCAACGCTTGGCTGTATCCCGGACGACCACTTC 1694
Qy 531 ArgThrAlaGlyTyPAspAlaLeuMetThrAsnProSerSerPhePheVal 547
Db 1695 CGCACCGCTGGCTACGACGCGTGTTCACCAACCGCTGCTGTTTGT 1745

RESULT 12
LOCUS CCLIP2 2043 bp DNA linear PLN 27-APR-1993
DEFINITION C.cylindracea LIP2 gene for lipase.
ACCESSION X64704
VERSION X64704.1 GI:2546
KEYWORDS Candida cylindracea
SOURCE Candida cylindracea
ORGANISM Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 2043)
AUTHORS Longhi,S., Fusetti,F., Grandori,R., Lotti,M., Vanoni,M. and
Alberghina,L.
TITLE Cloning and nucleotide sequences of two lipase genes from Candida
cylindracea
JOURNAL Biochim. Biophys. Acta 1131 (2), 227-232 (1992)
MEDLINE 92305068
PUBMED 1610906
REFERENCE 2 (bases 1 to 2043)
AUTHORS Longhi,S.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-1992) S. Longhi, Universita' degli Studi di
Milano, Via Celoria 26, 20133 Milano, ITALY
COMMENT See also x64703
x66006, x66007 and x66008 are related sequences in the description
of C.cylindracea Lip1 and Lip2 sequences (x64703, x64704).

FEATURES
Location/Qualifiers
1..2043
/organism="Candida cylindracea"
/mol_type="genomic DNA"
/strain="ATCC 14830"
/db_xref="taxon:44322"
/clone="B"
```

```
TATA_signal 126..132
gene 161..1807
/gene="LIP2"
CDS 161..1807
/gene="LIP2"
/product="triacylglycerol lipase"
/codon_start=1
/EC_number="3.1.1.3"
/protein_id="CAA45958.1"
/db_xref="GI:1325988"
/db_xref="GOA:P32946"
/translaton="MKUCLALGAARAAAPTATLANGDTITGLNALVNEKELGIPPEAE
PVPTLRFKPPVPSYASINGQQFTLYGLCPQMMPGSGFDTLPNNARHLVLQSKIFQ
VVLPNDSCLTINVIRPPTRASAGLPLMWIFGGFELGSSLPFGDQMVAKSVLMG
KPVTHVSMYRVASVWGLAGPDIQNEGSGAGLDORLQMWADNATAGFGDPSPKVT
IYGLAGSMSTFVHLVWNGDNTYKGPLPRAAIQMLGCMVPLDPDVTGTYFTIYDQV
VASGCSASDKLACGLGLDLYQATSDTPGVLAPESLRLVLPDPDGTGFIYDDMY
ALVRDGYARVFIIGQNDDEGLTGLLLNVTTDAARAYPKOLFHASDAEIDTLM
AAVTSIDITQGLPFDITGFNAITPQKRILALLGLDIAFLARRYELNYQGGTKYSELL
KQLLGLVPLGTGHNDITWQDYLVGSGSVIYNNAFIAPANDLDPNKAGLWNTWYITS
SLQGNLMLQINGLGLTYTGKDNFRPDAYSALFSNPPLFFV"
161..202
/gene="LIP2"
/EC_number="3.1.1.3"
sig_peptide 4.14e-176 Length: 2043
Pred. No.: 2212.00 Matches: 417
Score: 84.17% Conservative: 35
Percent Similarity: 77.65% Mismatches: 85
Best Local Similarity: 77.25% Indels: 0
Query Match: 8 Gaps: 0
ORIGIN
Alignment Scores:
Pred. No.: 4.14e-176 Length: 2043
Score: 2212.00 Matches: 417
Percent Similarity: 84.17% Conservative: 35
Best Local Similarity: 77.65% Mismatches: 85
Query Match: 77.25% Indels: 0
DB: 8 Gaps: 0
US-09-943-857-4 (1-547) x CCLIP2 (1-2043)
Qy 11 LeuGlySerValProThrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAla 30
Db 194 GTGGCGGACGACCCGCCACGCGCACCTCGCCACACGCGGACACCATCACCGCTCTCAACGCG 253
Qy 31 IleIleAsnGluAlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArg 50
Db 254 ATTGTCAACGAAAGTTTCTCGGCATACCGTTTCCGAGCGCCGCGGCGGCGCTCCG 313
Qy 51 PheLysAspProValProTyPheSerGlySerLeuAsnGlyGlnLysPheThrSerTyGly 70
Db 314 TTCAAGCGCGCGCTGCGGTACTCGCGCTCGCTCAACGCGGACGAGTTTACCCTGTACGCG 373
Qy 71 ProSerCysMetGlnGlnAsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAla 90
Db 374 CCGCTGTGATGCAGATGAACCTATGCGCTCGTTGAGGACACACTTCCCAAGATGCG 433
Qy 91 LeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCys 110
Db 434 CGGCATTTTGGTCTCCAGTCCAAGATCTTCCAAGTGTGCTTCCCAACGACGAGGACTGT 493
Qy 111 LeuThrIleAsnValValArgProProGlyThrLysAlaGlyAlaAsnLeuProValMet 130
Db 494 CTCACCATCAACGTGATCCGCGCGCGCGGACACGAGGCCAGTGCTGCTCCCGGTGATG 553
Qy 131 LeuTrpIlePheGlyGlyGlyPheGluIleGlySerProThrIlePheProAlaGln 150
Db 554 CTCGTGATCTTGGCGGTGGTTGAGCTTGGCGGCTCCAGGCTCTTCCAGGAGACGAC 613
Qy 151 MetValThrLysSerValLeuMetGlyLysHisIleIleHisValAlaValAsnTyArg 170
Db 614 ATGGTGGCCAAAGACGCTGCTCATGGGTAAACCGGCTGATCCACGTGAGCATGAACCTACCG 673
```

```
QY 171 ValAlaSerTrpGlyPheLeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsnAla 190
Db 674 GTGGCGTCATGGGGTCTTGGCCGGCCCCGACATCCAGAACGAGGCGAGCGAACGCC 733
QY 191 GlyLeuLysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGly 210
Db 734 GGCTTGATGACACGAGCGCTTGGCCATGAGTGGGTGGCGGACACATTTGCTGGGTGGC 793
QY 211 GlyAspProSerLysValThrIlePheGlyLysSerAlaGlySerMetSerValLeuCys 230
Db 794 GCGACCGGACGAGGTGACCATATACGCGAGCTGGCGGCGACATGTCAGCTTTGTG 853
QY 231 HisIleuIleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGly 250
Db 854 CACCTTGTGTGGACGACGCGGACGACACACGCTACAAACGCGAAGCGCTTGTTCGCGCGCC 913
QY 251 IleMetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGlu 270
Db 914 ATCATGACGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 973
QY 271 IleTyrAspLeuPheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCys 290
Db 974 ATCTACACCGAGTGTGGCTGTGCGGGGTGTGGAGTGCACGACGACAGCTGCGCTGC 1033
QY 291 LeuArgSerAlaSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeu 310
Db 1034 TTGGCGCGCTTCTGACGACGACCTTGTACAGCGCACGACGACGCGCGCGCTGTG 1093
QY 311 AlaTyrSerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThrAsp 330
Db 1094 GCGTACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1153
QY 331 AspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleIleGlyAsp 350
Db 1154 GACATGATGCTTGTGCGGACGCGAGTACGACACGCTGCGCGGTGATCATCGGCGAC 1213
QY 351 GlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnAlaGln 370
Db 1214 CAGAACGACGAGGCGACTTTGTTGGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1273
QY 371 AlaArgAlaTyrPheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeu 390
Db 1274 GCACGGCGGTACTTCAAGAGCTGTTTCAACGCGGACGATCGGAGATCGACACGTTG 1333
QY 391 MetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsn 410
Db 1334 ATGGCGCGGTACACGACGACATCACCGAGGCTGCGCTTGCACCGGCACTTCAAT 1393
QY 411 AlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHis 430
Db 1394 GCCATACCGCGAGTTCAAACGAGTCTCTGCGGTGCTTGGCGACCTTGGCGTTTCA 1453
QY 431 AlaArgTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSerLys 450
Db 1454 GCGGCTGCTACTTCTCAACTACTACCGGCGGACCAAGTACTCGTTTCTCCCTGAAG 1513
QY 451 GlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAsp 470
Db 1514 CAGCTTCTGGGTTGCGGCTTCTGGGACCTTCCACGGCGACGACATCATCTGGCAGGAC 1573
QY 471 TyrLeuLeuGlySerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAsp 490
Db 1574 TACTTGTGGGCGCGGCTGATCTACAAACGCGTTCATTGCGTTTGGCGTTCGACAGCAG 1633
QY 491 LeuAspProAsnThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerGln 510
Db 1634 CTGACCCGGAACAGGGGCGCTTGTGGACCAACTTGGCGCCACGTACACGACGCTGTCAG 1693
QY 511 SerGlyAsnAsnLeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPhe 530
Db 1694 CTGGGCAACACTTGATGACAGTCAACGGCTTGGGTTGTACACCGGCAAGGACCACTTC 1753
QY 531 ArgThrAlaGlyTyrAspAlaLeuMetThrAsnProSerSerPhePheVal 547
```

```
Db 1754 CGCCCGGATCGGTACAGCGCCTCTTTTCCAAACCGCCACTGTTCTTTGTG 1804
RESULT 13
AX709925
LOCUS
DEFINITION
Sequence 7 from Patent EPI288294.
AX709925
ACCESSION
AX709925.1 GI:29786304
KEYWORDS
SOURCE
ORGANISM
Candida rugosa
Candida rugosa
Bukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE
1
Tang, S. J., Lee, G. C. and Shaw, J. F.
Recombinant candida rugosa lipases
Patent: EP 1288294-A 7 05-MAR-2003;
Academia Sinica (TW)
AUTHORS
Tang, S. J., Lee, G. C. and Shaw, J. F.
TITLE
Recombinant candida rugosa lipases
JOURNAL
Academia Sinica (TW)
FEATURES
Location/Qualifiers
source
1..1541
/organism="Candida rugosa"
/mol_type="unassigned DNA"
/db_xref="taxon:5481"
ORIGIN
Alignment Scores:
Pred. No.: 2,53e-161 Length: 1541
Score: 2034.00 Matches: 434
Percent Similarity: 87.57% Conservative: 31
Best Local Similarity: 81.73% Mismatches: 49
Query Match: 71.02% Indels: 24
DB: 6 Gaps: 15
US-09-943-857-4 (1-547) x AX709925 (1-1541)
QY 17 AlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAlaIleAsnGluAlaPhe 36
Db 7 GCCACGCTCGCCAAACGGGACACCATCACCGGTCTCAACGCCATCATCAACGAGGCTTC 66
QY 37 LeuGlyIleProPheAlaGluProValGlyAsnLeuArgPheLysAspProValPro 56
Db 67 CTCGGCATTCCTTTTGGGAGCGCGGTGGGCAACCTCCGCTTCAAGGACCCCTGTGCG 126
QY 57 TyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyProSerCysMetGlnGln 76
Db 127 TACGTGGGTCTCTCAACGGTCAATCTTCCCGGCTACGCTCCG---TGCATGCACGAG 183
QY 77 AsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGln 96
Db 184 AACCCCGAGGCGACCTACGAGGAGAACCTCCCAAGGTGGCGCTTGACTTGGTATGTCAG 243
QY 97 SerLysValPheGlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsnValVal 116
Db 244 TCCAAGGTGTTCAGGCTGTCTCCGCCAACAGGAGACTGSCCTCACCATCAACGCTGGTG 303
QY 117 ArgProGlyThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyGly 136
Db 304 CGCCCGCGGCGACCAAGGCGGCGCAACCTCCCGGTATGCTCTGGATCTTTGGCGGT 363
QY 137 GlyPheGluIleGlySerProThrIlePheProAlaGlnMetValThrLysSerVal 156
Db 364 GGGTTTGAGTACGACGCGCCACCATCTTCCCTCCCGCTCAGATGCTCCAGAGTGTG 423
QY 157 LeuMetGlyLysHisIleIleHisValAlaValAsnTyrArgValAlaSerTrpGlyPhe 176
Db 424 CTCAATGGGAGGCC-ATCATCCAGCTGGCGCTCAACTACCGCTTGGCTCTTGTGTTTC 482
QY 177 LeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArg 196
Db 483 TTGGCGGTTCCGAGACATCAAGCGGAGGCGAGCTCCCAATGCCCGGCTCAAGGACCGC 542
QY 197 LeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysVal 216
```

Db	543	TTGGGATGAGTGGTGGCAGACCACTTCGCGGGTTGGCGGACCCGAGCAAGGTG	602		AX709923	1548 bp	DNA	linear	PAT 10-APR-2003
Qy	217	ThrIlePheGlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAsp	236		SEQUENCE	5 from Patent EP1288294.			
Db	603	ACCATCTTTGGGAG---GCGGGCAGCATGTCCGTGTGTGGCCACCTTCTCTGGAATGGC	658		ACCESSION	AX709923			
Qy	237	GlyAspAnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAla	256		VERSION	AX709923.1	GI:29786303		
Db	659	GGCGACAAACAGTCAAGGGCAAGCGTGTTCGCGGGCATCATGACG---GGAGCC	715		KEYWORDS	Candida rugosa			
Qy	257	MetValProSerAspProValAspGlyThrTyrGlyAsnGluIleTyrAspLeuPheVal	276		ORGANISM	Candida rugosa			
Db	716	ATGGTGCGG---GACCCCGTGGACGGCACTATGGACC-CAAAATCATGACAGTGTGGT	771		REFERENCE	Tang, S.J., Lee, G.C. and Shaw, J.F.			
Qy	277	SerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCysLeuArgSerAlaSerSer	296		AUTHORS	Recombinant candida rugosa lipases			
Db	772	GCTACG---GGCTGCAGCAGTGCCAGCAACAGCTTGGCTGCTTCGCTGGTCTT---ACT	825		TITLE	Patent: EP 1288294-A 5 05-MAR-2003;			
Qy	297	AspThrLeuLeuAspAlaThrAsnAnThrProGlyPheLeuAlaTyrSerSerLeuArg	316		JOURNAL	Academia Sinica (TW)			
Db	826	CAGGCATTGCTCGATGCCACCAACGACACCCCTGGGTTCTTGTCGTACACCTCGTTGGG	885		FEATURES	Location/Qualifiers			
Qy	317	LeuSerTyrLeuProArgProAspGlyLysAnlleThrAspAspMetTyrLysLeuVal	336		source	1. 1548			
Db	886	TTG-----TACTCCCGGCCCGACGCGCCACATCCGATGACATGTACAAAGTTGGTA	939		/organism="Candida rugosa"				
Qy	337	ArgAspGlyLysTyrAlaSerValProValIleIleGlyAspGlnAsnAspGluGlyThr	356		/mol_type="unassigned DNA"				
Db	940	CGCGACGGCAAGTATGCAAGCGTTCCCGTGTATCAVTGGCGACCAAGACGAGGGCTTC	999		/db_xref="taxon:5481"				
Qy	357	IlePheGlyLeuSerSerLeuAnValThrThrAsnAlaGlnAlaArgAlaTyrPheLys	376		ORIGIN				
Db	1000	TTGTTTGT-CTC-----TTGAACACCACCGAGCGCGGACGATACCTCAGA	1052		Alignment Scores:				
Qy	377	GlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyrProGln	396		Pred. No.:	3,82e-160	Length:	1548	
Db	1053	AAG---TTCATCCAGCCACCGACCGGATATCACCGCATTCAGCGCGGTACCCACGC	1109		Score:	2020.00	Matches:	422	
Qy	397	AspIleThrGlnGlySerProPheAspThrGlyValLeuAsnAlaLeuThrProGlnPhe	416		Percent Similarity:	87.12%	Conservative:	38	
Db	1110	GATGTCAACCCAGGGT---CCGTTCGACACGGGCATTTCTCAACGCCCTTACACCCAGCTC	1166		Best Local Similarity:	79.92%	Mismatches:	57	
Qy	417	LysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHisAlaArgArgTyrPheLeu	436		Query Match:	70.53%	Indels:	17	
Db	1167	AAGCGATCAATGCTGTGCTTGGCGACCTCACTTTACCTTCGCGCGGCTACTTCCTC	1226		DB:	6	Gaps:	9	
Qy	437	AsnHisPheGlnGlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeuPro	456		US-09-943-857-4 (1-547) x AX709923 (1-1548)				
Db	1227	AACCACTACACCGGTGTCCTCAAGTACTCGTTCTCTC---AAGCAGCTT---GGTTG-CCG	1279		Qy	15	ProthrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAlaIleIleAsnGlu	34	
Qy	457	IleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAspTyrLeuLeuGlySerGly	476		Db	2	CCCCTGCCAGCTCGCCCAACGGCGACACCATCACCGGTCTCAACGCCATCATCAACGAG	61	
Db	1280	ATTCTCGGACG-TTCACCGCAACGACATGTGTGGCAGACATTTTGTGGCGACGCGC	1338		Qy	35	AlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArgPheLysAspPro	54	
Qy	477	SerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAnThrAla	496		Db	62	CGCTTCCCTCGGTATTTCCTTGTCTCAGCGCGCGGTGGGCACTCCCGCTTCAAGCGCCT	121	
Db	1339	AGCGTCACTCAACACGCGTTTCATCGGTTTGCACCGACTTGGACCCCAACACCGCG	1398		Qy	55	ValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyProSerCysMet	74	
Qy	497	GlyLeuLeuValAsnTrpProLysTyrThrSerSerGlnSerSerGlyAsnAnLeuMet	516		Db	122	GTGCGGTACTCGGCGTCTCTCAATGGTCAGAAGTTTACTGTATGG---CCCTG-TGCAATG	177	
Db	1399	GGC---TTGGTGCAGTGGCCCAAGTC-ACC---AGCAGCGAGCGCGGAGACTGTCTCACC	1451		Qy	75	GlnGlnAnProGluGlyThrPheGluGluAnLeuGlyLysThrAlaLeuAspLeuVal	94	
Qy	517	MetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAnPheArgThrAlaGlyTyrAsp	536		Db	178	CAGATGAACCCATTGGGCAACTTGGGACTCTCGCTTCCCAAGGCTGCCATCAACTTG---	234	
Db	1452	CAGATCAGTGCCTTGGGCTGTGTACACCGGCAAGGACAACTTCGCGACCCCGGCTACAAC	1511		Qy	95	MetGlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsn	114	
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547		Db	235	ATGCAGTCCAGCTCTTCAGCGGTGTCTTAAACGGCGAGGACTGTCTCACCATCAAC	294	
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541		Qy	115	ValValArgProProGlyThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePhe	134	
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547		Db	295	GTGGTGCGG---CCGCGCACCAAGCGGGGTGCCAACCTCCCGGTGATGTGTGGAITTTT	351	
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541		Qy	135	GlyGlyGlyPheGluLeuGlySerProThrIlePheProProAlaGlnMetValThrLys	154	
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547		Db	352	GGCGCGGGTTTGAGGTGGCGGCTCCAGTCTCTTCCCTCCCGCAACAGATGATCACCGCC	411	
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541		Qy	155	SerValLeuMetGlyLysHisIleIleHisValAlaValAsnTyrArgValAlaSerTrp	174	
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547		Db	412	AGCGTGTATTGGGCAAGCCCATATCCACGTGAGCATACCGGCTTGTTCGTGG	471	
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541		Qy	175	GlyPheLeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAsp	194	
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547		Db	472	GGGTCTTGGCTGGTCCAGACATCAAGCCGAGGGCAGCGGAAACCCCGGTTTGACAGAC	531	
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541		Qy	195	GlnArgLeuGlyMetGlnTrpValAlaAspAnlleAlaGlyPheGlyGlyAspProSer	214	
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547		Db	532	CAACGCTTGGGTGTGAGTGGGTGGCGGACAACTTGGCGGGTTCCGCGGACCCCGTCC	591	
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541		Qy	215	LysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrp	234	
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						
Db	1512	GCTTGTGTCGCGACCCG---CACTTTTTCGTG	1541						
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547						

Db 592 AAGGTGACCATCTTTGGTGAGG--CGCGGCGACATGTCGGTAATGTCAGCTCTCTGG 649

QY 235 AsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyLeuMetGlnSer 254

Db 650 AACGACGCGGACACACGTCACACGCGAAGCGGTGTTCCGTGCGCGCATCATGCAG--- 706

QY 255 GlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGluLeuTyrAspLeu 274

Db 707 GGGGCCCATGGTGGCGG--GACCCGGTGGATGGCGCTACGGCAGCAGCATCTAGCACCG 764

QY 275 PheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCysLeuArgSerAla 294

Db 765 GTGGTTGCTTCAGCGCGGTGTGGAGTCCAGCAGCAAGCTCGCGGTGTTCCGCA--GCA 822

QY 295 SerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeuAlaTyrSerSer 314

Db 823 TCGAACGACAACTCTTCCAGGCCACCGACGACACTCGGGGGCTTGGCGTACCCCTCG 882

QY 315 LeuArgLeuSerTyrLeuProArgProAspGlyLysAlaThrAspAspMetTyrLys 334

Db 883 TTGCGGTGG--TTTCTCCCGCGCGCCGACGCGACCTTTCATCAGCATGATGTTCAAG 940

QY 335 LeuValArgAspGlyLysTyrAlaSerValProValLeuLeuGlyAspGlnAsnAspGlu 354

Db 941 TTGGTGGCGGCGGCAAGTGTCCACAGTTCCGGTGATCATTTGGCGACCAAGACGAG 1000

QY 355 GlyThrLeuPheGlyLeuSerSerLeuAsnValThrThrAsnAlaGlnAlaArgAlaTyr 374

Db 1001 GGCACAGTGTTCGCTTG-----TTGAACGTGACTACGGATGCTCAGGACGCGCATC 1054

QY 375 PheLysGlnSerPheLeuHisAlaSerAspAlaGluLeuAspThrLeuMetAlaAlaTyr 394

Db 1055 TTCAAGGAA--TTTCAATCCACGCGCAGCGAGATGACACCTTTGATGGCGCGTAC 1111

QY 395 ProGlnAspLeuThrGlnGlySerProPheAspThrGlyValLeuAsnAlaLeuThrPro 414

Db 1112 CCAGCAGCATATCCACAGGTT---CGTTTCGACACCGCGCATCTTCAACGCGCATCCCG 1168

QY 415 GlnPheLysArgLeuSerAlaValLeuGlyAspLeuAlaPheLeuHisAlaArgArgTyr 434

Db 1169 CAGTTCAACCGGATTCAGCGGTGCTTGGTGACCTTTCGTTCTCTCCCGCGCGTAC 1228

QY 435 PheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSerLysGlnLeuSerGly 454

Db 1229 TTCTCTCAACCATCTCCAGGCGCGCACCAAGTACTCGTTCC--TCGAAGCAGCTT---GGG 1283

QY 455 LeuProLeuMetGlyThrPheHisAlaAsnAspLeuValThrGlnAspTyrLeuLeuGly 474

Db 1284 TTGCGCGTATTCGACCCACACGCGCATTCGTTGACCTTTCGTTCTCTCCCGCGCGTAC 1343

QY 475 SerGlySerValLeuTyrAsnAsnAlaPheLeuAlaPheAlaThrAspLeuAspProAsn 494

Db 1344 CACAGCAGCGCGGTGTACAAACGCGTTTATTGGCTTTTGCCAAACGACCTCGACCCGAC 1403

QY 495 ThrAlaGlyLeuLeuValAsnThrProLysTyrThrSerSerSerGlnSerGlyAsnAsn 514

Db 1404 AAGCGCGGTGTGTTGTAAGTGGCCCAAGTATACACGAGCAGC-----CAGGGCAAC 1457

QY 515 LeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGly 534

Db 1458 TTGTTGCAGTCAACGCTTGGGCTGTATACCGCGCAAGACAACTTCGCGACCGCTGGC 1517

QY 535 TyrAspAlaLeuMetThrAsnPro 542

Db 1518 TAGCAGCGGTGTTTACCAACCG 1541

RESULT 15

AX709927

LOCUS

DEFINITION

Sequence 9 from Patent EP1288294.

ACCESSION

AX709927

VERSION

AX709927.1

KEYWORDS

GI:29786305

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-09-943-857-4 (1-547) x AX709927 (1-1511)

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

Candida rugosa

Candida rugosa

Saccharomycetales; mitosporic Saccharomycetales; Candida.

Tang, S.J., Lee, G.C. and Shaw, J.F.

Recombinant candida rugosa lipases

Patent: EP 1288294-A 9 05-MAR-2003;

Academia Sinica (TW)

Location/Qualifiers

1..1511

/organism="Candida rugosa"

/mol_type="unassigned DNA"

/db_xref="taxon:5481"

2,49e-150

1903.00

86.84%

81.39%

66.45%

6

29

43

38

20

1511

433

29

43

38

20

1511

433

29

43

38

20

1511

433

29

43

38

20

1511

433

29

43

38

20

1511

433

29

43

38

20

1511

433

29

43

38

20

1511

433

29

43

38

20

1511

433

29

43

38

20

1511

433

29

43

38

20

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 5, 2004, 17:57:31 ; Search time 449 Seconds

(without alignments)
5175.422 Million cell updates/sec

Title: US-09-943-857-4

Perfect score: 2864

Sequence: 1 SMNSRGPAGRLGSLVPTAKLA.....DNFRTAGYDALMTNPSSFFV 547

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlh
-O=/cgn2.1/USPTO.spool/US09943857/runat.29072004.090859.29721/app.query.fasta_1.711
-DB=N Geneseq 29Jan04 QFMT=fastap -SUFFIX=ing -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptto -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09943857 @CGN 1.1.352 @runat.29072004.090859.29721 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N Geneseq 29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2864	100.0	1641	7 ABX95906	Abx95906 Candida r
2	2588	90.4	1641	7 ABX95909	Abx95909 Candida r
3	2513	87.7	1647	2 AAX33112	Aax33112 Synthetic
4	2512	87.7	1950	4 AAH43625	Aah43625 Synthetic
5	2495	87.1	1641	7 ABX95908	Abx95908 Candida r
6	2487	86.8	1649	2 AAX33111	Aax33111 Candida r
7	2434	85.0	1641	7 ABX95907	Abx95907 Candida r
8	2412	84.2	1650	2 AAT10422	Aat10422 Candida c

9	2396	83.7	1641	7 ABX95905	Abx95905 Candida r
10	1107	38.7	1828	2 AAQ54020	Aaq54020 Lipase co
11	1097	38.3	1692	2 AAQ10313	Aaq10313 Sequence
12	1088	38.0	1635	2 AAT10421	Aat10421 Geotrichu
13	1078	37.6	1674	2 AAQ05605	Aaq05605 Gene enco
14	938.5	32.8	2045	2 AAQ98578	Aaq98578 Aspergill
15	938.5	32.8	2045	2 AAT11299	Aat11299 Aspergill
16	632	22.1	8528	2 AAQ46249	Aaq46249 Construct
17	481	16.8	2089	6 ABI99790	Abi99790 Mouse isc
18	467	16.3	2544	7 ABZ81313	Abz81313 Human dru
19	462	16.1	1698	2 AAX86613	Aax86613 cDNA enco
20	462	16.1	2478	2 AAX86614	Aax86614 cDNA enco
21	452.5	15.8	1738	2 AAV21460	Aav21460 Rhodospo
22	452	15.8	1725	6 AAS17549	Aas17549 Synthetic
23	452	15.8	1752	3 ABA97180	Aba97180 Ache codi
24	452	15.8	1845	6 AAS17493	Aas17493 Human cdn
25	452	15.8	2156	7 ABZ23128	Abz23128 Nucleotid
26	452	15.8	2256	2 AAQ99002	Aaq99002 Human ace
27	452	15.8	2400	2 AAQ42496	Aaq42496 Full-leng
28	452	15.8	5767	6 AAS17547	Aas17547 Plasmid p
29	452	15.8	14446	6 AAS17548	Aas17548 Plasmid p
30	451	15.7	1725	8 ACC84169	Acc84169 Human but
31	451	15.7	1809	8 ACC84170	Acc84170 Human but
32	451	15.7	2381	6 ABN95773	Abn95773 Gene #227
33	451	15.7	2416	3 AAZ49471	Aaz49471 Human but
34	451	15.7	2416	3 AAZ49470	Aaz49470 Human wil
35	451	15.7	2416	6 AAL49277	Aal49277 Human but
36	451	15.7	2444	6 ABK32867	Abk32867 DNA encod
37	449.5	15.7	1716	4 AAH26618	Aah26618 Rhodospo
38	449.5	15.7	1967	6 AAL49276	Aal49276 Human but
39	444.5	15.5	2445	1 AAN60111	Aan60111 Sequence
40	444.5	15.5	2445	2 AAQ48393	Aaq48393 Full-leng
41	438	15.3	2424	8 ACH03894	Ach03894 Human cdn
42	438	15.3	2520	9 ADD12610	Add12610 Human ENZ
43	437.5	15.3	1470	2 AAT76597	Aat76597 Modified
44	437.5	15.3	1470	2 AAX80854	Aax80854 Para-Nitr
45	436.5	15.2	1470	2 AAX80851	Aax80851 Para-Nitr

RESULT 1

ABX95906

ID ABX95906 standard; DNA; 1641 BP.

XX

AC ABX95906;

XX

DT 15-JUL-2003 (first entry)

XX

DE Candida rugosa lipase 3 DNA.

XX

DE Lipase 3; gene; ds; mutant.

XX

OS Candida rugosa.

XX

Key Location/Qualifiers

FT CDS 1..1641

FT /tag= a

FT /product= "Lipase 3"

FT /partial

FT /note= "No start or stop codon shown"

XX

PN EP1288294-A2.

XX

PD 05-MAR-2003.

XX

PF 26-APR-2002; 2002EP-00009616.

XX

PR 31-AUG-2001; 2001US-00943857.

XX

PA (SINI-) ACAD SINICA.

XX

PI Tang S, Lee G, Shaw J;

ALIGNMENTS

XX WPI; 2003-395476/38.
 DR P-PSDB; ABU09071.
 XX
 PT Isolated mutant nucleic acid encoding Candida rugosa lipase, useful for
 the preparation of Candida rugosa lipase for biocatalytic applications.
 XX
 PS Claim 25; Page 5-7; 33pp; English.
 XX
 CC The invention relates to an isolated mutant nucleic acid encoding a
 CC Candida rugosa lipase polypeptide. The DNA has a sequence having at least
 CC 80% identity to a wild-type DNA encoding Candida rugosa lipase. The
 CC invention also relates to a microorganism comprising the DNA, where the
 CC microorganism is a bacterium or yeast, preparing a mutant DNA encoding a
 CC C.rugosa lipase and a chimeric C.rugosa lipase comprising a substrate
 CC interacting domain of a first C.rugosa lipase and a non-substrate
 CC interacting domain of a second C.rugosa lipase. The method is useful for
 CC preparing a mutant DNA encoding a Candida rugosa lipase. The DNA is
 CC useful in the large scale manufacture of Candida rugosa lipase which is
 CC useful for biocatalytic applications. This sequence represents DNA
 CC encoding Candida rugosa lipase 3
 XX
 SQ Sequence 1641 BP; 315 A; 525 C; 459 G; 342 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 3 22e-263 Length: 1641
 Score: 2864.00 Matches: 547
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 7 Gaps: 0
 US-09-943-857-4 (1-547) x ABX95906 (1-1641)
 QY 1 SerMetAsnSerArgGlyProAlaGlyArgLeuGlySerValProThrAlaLysLeuAla 20
 DB 1 TCGATGAATTCACGTGGCCAGCGCGCTCCGATCGGTACCCACCGCCAAAGTCGCG 60
 QY 21 AsnGlyAspThrIleThrGlyLeuAsnAlaIleAsnGluAlaPheLeuGlyIlePro 40
 DB 61 AACGGCGACACATCACCGGTCTCAACGCCATCATCAAGAGCGGTTCCTCGCATTTCC 120
 QY 41 PheAlaGluProValGlyAsnLeuArgPheLysAspProValProThrSerGlySer 60
 DB 121 TTTCGGACCGCGCGGTGGGCAACCTCCGCTTCAAGACCTGTGCGTACTTGGCTCG 180
 QY 61 LeuAsnGlyGlnLysPheThrSerThrGlyProSerCysMetGlnGlnAsnProGluGly 80
 DB 181 CTCACGGCCAGAAAGTTCACTTCTTACGGCCGCTTTCATGCAGCAACCCCGAGGCG 240
 QY 81 ThrPheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPhe 100
 DB 241 ACGTTTGAAGAGAACCTTGGCAAGACGGCAGCTCGACTTGGTGATGAGTCCAGGTTTC 300
 QY 101 GlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgProProGly 120
 DB 301 CAGGCGGTGCTTCCCGAGTGGAGTGCCTCACCATCAACGTGGTGGCGCGCGGCG 360
 QY 121 ThrLysAlaGlyAlaAsnLeuProValMetLeuThrIlePheGlyGlyGlyPheGluIle 140
 DB 361 ACCAAGCGGGGCCAACCTCCCGGTCTATGCTCTGATCTTTGGCGGTGGTGTGATC 420
 QY 141 GlySerProThrIlePheProProAlaGlnMetValThrLysSerValLeuMetGlyLys 160
 DB 421 GGCAGCGCCACCATCTTCCCTCCCGCCAGATGGTCACCAAGAGTGTGCTCATGGGCAAG 480
 QY 161 HisIleIleHisValAlaValAsnThrArgValAlaSerTrpGlyPheLeuAlaGlyAsp 180
 DB 481 CACATCATCACTGGCGGTCACTACCGGTGTTCCTCGTGGGGTTCCTGGCTGTGAT 540
 QY 181 AspIleLysAlaGluGlySerGlyAsnAlaGlyLysAspGlnArgLeuGlyMetGln 200
 DB 541 GACATCAAGGCCGAGGCGAGCGGAAAGCGCGGTGGAAGACAGCGTTTGGGCGATGCAG 600

QY 201 TrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysValThrIlePheGly 220
 DB 601 TGGGTGGCAGACAACATTGGCGGTTCGGCGGCGACCCGAGCAAGGTGACTATCTTTGGC 660
 QY 221 GluSerAlaGlySerMetSerValLeuCyHisIleuIleTrpAsnAspGlyAspAsnThr 240
 DB 661 GAGTCTCGGGCAGCATGTCGTGTGTGGCCACCTCATCTGGAACGACGGCGCAACACG 720
 QY 241 TyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValProSer 260
 DB 721 TACAAAGGCGAAGCGTGTTCGGCGGGGATCATGAGTCTGGAGCATGGTGGCGTCT 780
 QY 261 AspProValAspGlyThrTyrGlyAsnGluIleTyrAspLeuPheValSerSerAlaGly 280
 DB 781 GACCGGTGGAGCGCACGTACCGCAACGAGATCTACGACCTCTTTGTCTCGAGTGTGGC 840
 QY 281 CysGlySerAlaSerAspLysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeu 300
 DB 841 TGTGGCAGCGCAGCGCAAGCTCGCGTGTTCGGCAGTGGCTTAGCGACACCTTGTGTC 900
 QY 301 AspAlaThrAsnAsnThrProGlyPheLeuAlaTyrSerSerLeuArgLeuSerTyrLeu 320
 DB 901 GATGCCACCAACACATCTCTGGGTCTTGGGTACTCTCTGTTGGGTGTCTTTATCTC 960
 QY 321 ProArgProAspGlyLysAsnIleThrAspAspMetTyrLysLeuValArgAspGlyLys 340
 DB 961 CGCGGCGCGCAGCGCAACATCACCGATGACATGTACAAGTTTGGTGGCGACGGCAAG 1020
 QY 341 TyrAlaSerValProValIleIleGlyAspGlnAsnAspGluGlyThrIlePheGlyLeu 360
 DB 1021 TATGCAAGCGTTCGCGTATCATTTGGCAGCAACGACGAGGCGACCATCTTTGGGCTC 1080
 QY 361 SerSerLeuAsnValThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPheIle 380
 DB 1081 TCTTTTGTAAAGTGGACCAAGATGCTCAGGCGCGTGTACTTCAAGCAGTCTTTATC 1140
 QY 381 HisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyrProGlnAspIleThrGln 400
 DB 1141 CAGCGCAGCGCAGCGAGATCGACACTTGTATGGCGGGGTACCCCGCAGGACATCACCAG 1200
 QY 401 GlySerProPheAspThrGlyValLeuAsnAlaLeuThrProGlnPheLysArgIleSer 420
 DB 1201 GGTTCCTCGGTTCGACACCGGTGTCTCAACGCGCTTCAACCGCGCTTCAAGAGAATCTCT 1260
 QY 421 AlaValLeuGlyAspLeuAlaPheIleHisAlaArgArgTyrPheLeuAsnHisPheGln 440
 DB 1261 GCGGTGCTCGGCGACCTTGCATTCATCCACGCGCGCGCTACTTCTCAACCATTCACG 1320
 QY 441 GlyGlyThrLysThrSerPheLeuSerLysGlnLeuSerGlyLeuProIleMetGlyThr 460
 DB 1321 GCGCGCACCAAGTACTCGTTCCTCTTAAGCAGCTCTCTGGGTTCGCAATCATGGGCACC 1380
 QY 461 PheHisAlaAsnAspIleValTrpGlnAspTyrLeuLeuGlySerGlySerValIleTyr 480
 DB 1381 TTTCATGCGCAACGACATGTGTGGCAGGACTACTTGTGGGAAGCGCGAGCTCATCTAC 1440
 QY 481 AsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAlaGlyLeuLeuVal 500
 DB 1441 AACACGCGTTCATCGGTTCGCCACCGACTTGGACCCCAACACCGCGGGTGTGGTGTG 1500
 QY 501 AsnTrpProLysThrSerSerSerGlnSerGlyAsnAsnLeuMetMetIleAsnAla 520
 DB 1501 AACTGGCCCAAGTACACGACGAGCTCTAGTCTGGCAACAACTTGTATGATGATCAACGCG 1560
 QY 521 LeuGlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGlyTyrAspAlaLeuMetThr 540
 DB 1561 TTGGGCTTGTACACCGGCAAGGACAACTTCCGCAACGCTGGCTAGCAGCGGTGTGATGACC 1620
 QY 541 AsnProSerSerPhePheVal 547
 DB 1621 AACCGTCTTCTTCTTCTGTG 1641

RESULT 2

ABX95909
ID ABX95909 standard; DNA; 1641 BP.

XX AC ABX95909;

XX DT 15-JUL-2003 (first entry)

XX DE Candida rugosa lipase 8 DNA.

XX KW Lipase 8; gene; ds; mutant.

XX OS Candida rugosa.

XX FH Key Location/Qualifiers
FT CDS 1. .1641
FT FT /*tag= a
FT FT /product= "Lipase 8"
FT FT /partial
FT FT /note= "No start or stop codon shown"

XX PN EP1288294-A2.

XX XX 05-MAR-2003.

XX XX 26-APR-2002; 2002EP-00009616.

XX XX 31-AUG-2001; 2001US-00943857.

XX PA (SINI-) ACAD SINICA.

XX PI Tang S, Lee G, Shaw J;

XX XX WPI; 2003-395476/38.

DR P-PSDB; ABU09074.

XX PT Isolated mutant nucleic acid encoding Candida rugosa lipase, useful for
the preparation of Candida rugosa lipase for biocatalytic applications.

XX PS Claim 25; Page 10-11; 33pp; English.

XX CC The invention relates to an isolated mutant nucleic acid encoding a
Candida rugosa lipase polypeptide. The DNA has a sequence having at least
80% identity to a wild-type DNA encoding Candida rugosa lipase. The
invention also relates to a microorganism comprising the DNA, where the
microorganism is a bacterium or yeast, preparing a mutant DNA encoding a
C. rugosa lipase and a chimeric C. rugosa lipase comprising a substrate
interacting domain of a first C. rugosa lipase and a non-substrate
interacting domain of a second C. rugosa lipase. The method is useful for
preparing a mutant DNA encoding a Candida rugosa lipase. The DNA is
useful in the large scale manufacture of Candida rugosa lipase which is
useful for biocatalytic applications. This sequence represents DNA
encoding Candida rugosa lipase 8

XX SQ Sequence 1641 BP; 298 A; 537 C; 472 G; 334 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 6,89e-237 Length: 1641
Score: 2588.00 Matches: 485
Percent Similarity: 94.15% Conservative: 30
Best Local Similarity: 88.67% Mismatches: 32
Query Match: 90.36% Indels: 0
DB: 7 Gaps: 0

US-09-943-857-4 (1-547) x ABX95909 (1-1641)

QY 1 SerMetAsnSerArgGlyProAlaGlyArgLeuGlySerValProThrAlaLysLeuAla 20

DB 1 TCGATGAATTCAGTGGCGCAGCGCGCGTCTCGGATCGGTACCCACTGCCAGCTCGCC 60

QY 21 AsnGlyAspThrIleThrGlyLeuAnAlaIleAsnGluAlaPheLeuGlyIlePro 40

DB 61 AACGGCGACCATCATCCGGTCTCAACGCCCATCATCAACGAGCGGTCTCTCGGCATTCCC 120

QY 41 PheAlaGluProValGlyAsnLeuArgPheLysAspProValProTyrSerGlySer 60
DB 121 TTTGCCGAGCGCGCGGTGGGCAACCTCCGCTTCAAGGACCCCGTACTCTCGGCTCG 180
QY 61 LeuAsnGlyGlnLysPheThrSerTyrGlyProSerCysMetGlnGlnAsnProGluGly 80
DB 181 CTCGATGCCAGAGTTCACTTCTTACGGCCCGCTTGCATGCAGCAGAACCCCGAGGCG 240
QY 81 ThrPheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPhe 100
DB 241 ACCTACGAGGAGAACCTCCCAAGCGCAGCTCGACTTGGTGTGTCAGTCCCAAGTGTGT 300
QY 101 GlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgProGly 120
DB 301 GAGCGGTGTCTCGTCTAGCGAGACTGTCTCACCATCAACGCTGGTGGCGCCCGCGGC 360
QY 121 ThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyGlyGlyPheGluIle 140
DB 361 ACCAAGCGGGTGCCAACTCCCGGTGTCTGTGATCTTTGGCGGGGGTTGAGGTG 420
QY 141 GlySerProThrIlePheProAlaGlnMetValThrLysSerValLeuMetGlyLys 160
DB 421 GGTGGCACCAACCTTCCCTCCCGCCAGATGATCACAAGAGCATTTGCCATGGCAAG 480
QY 161 HisIleIleHisValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGlyAsp 180
DB 481 CCCATCATCCAGTGGAGCTCAACTACCGCTGTCTGGGGGTCTTTGGTGGCGCAC 540
QY 181 AspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMetGln 200
DB 541 GAGATCAAGGCGCGAGGCGAGTGCACACGCGGTTCAGAGGACCCAGCGCATGGCGATGCAG 600
QY 201 TrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysValThrIlePheGly 220
DB 601 TGGGTGGCGGACAACTTGGCGGTGGCGGCGACCGACCAAGGTGACCATCTTTGGC 660
QY 221 GluSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAspGlyAspAsnThr 240
DB 661 GAGTCTGGCGGCGAGCATGTCGTATGTCACATCTCTGGAACACGCGCGCAACACG 720
QY 241 TyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValProSer 260
DB 721 TACAAGGCGCAAGCGCTCTTCGCGCGGCGCATCATGCAGTCTGGGCGCATGTGACGTGC 780
QY 261 AspProValAspGlyThrTyrGlyAsnGluIleTyrAspLeuPheValSerSerAlaGly 280
DB 781 GAGCGGTGGACGCGCTCTACGCGCAACGAGATCTTTGACCTCTGGCGTCGCGACGCGGC 840
QY 281 CysGlySerAlaSerAspLysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeu 300
DB 841 TCGCGGCGCGCGACGCGACGACGCTTGGCTGTCTGGCGGGTGTGTAGCGACACGTTGGAG 900
QY 301 AspAlaThrAsnAsnThrProGlyPheLeuAlaTyrSerSerLeuArgLeuSerTyrLeu 320
DB 901 GAGCGCACCAACACACCCCTGGGTCTTGGCGTACTCTCTCGTTCGGGTGTCTTATCTC 960
QY 321 ProArgProAspGlyLysAsnIleThrAspAspMetTyrLysLeuValArgAspGlyLys 340
DB 961 CGCGCGCGCGCGGTGTAACATCACCGACGACATGTTTGGCTTGGTCCGCGAGGCGCAAG 1020
QY 341 TyrAlaSerValProValIleIleGlyAspGlnAsnAspGluGlyThrIlePheGlyLeu 360
DB 1021 TATGAAGCGTTCCTGTGTATCATCGCGCACCAAGACGAGCGGCGCACCTTCTTGGCAC 1080
QY 361 SerSerLeuAsnValThrThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPheIle 380
DB 1081 TCTTCTTTGAACGTGACCAAGATCGCGAGGCGCGCGAGTACTTTCAGCAGTCTTTTGTG 1140
QY 381 HisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyrProGlnAspIleThrGln 400
DB 1141 CACGCGCAGCGCGAGCTCGACACGTTGATGAGCGGTGTACCCCGAGGACATCACCCAG 1200

Fri Aug 6 10:53:11 2004

us-09-943-857-4.rng

QY 401 GlySerProPheAspThrGlyValLeuAsnAlaLeuThrProGlnPheLysArgIleSer 420
 DB 1201 GGTTCCTCGTTCGACACGGGTGTCTCAACGCCCTCACCCCGCAGTTTCAAGAGAACTCT 1260
 QY 421 AlaValLeuGlyAspLeuAlaPheIleHisAlaArgArgTyrPheLeuAsnHisPheGln 440
 DB 1261 GCGGTGCTCGCGACCTTGTCTTCATCCACGCCGCTGCTACTTCTCTCAACCACTACACC 1320
 QY 441 GlyGlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeuProIleMetGlyThr 460
 DB 1321 GCGGACCAAGTACTCATCTCTTAAGCAGCTCTCTGGCTTGGCGGTGCTCGAAGC 1380
 QY 461 PheHisAlaAsnAspIleValTrpGlnAspTyrLeuLeuGlySerGlySerValIleTyr 480
 DB 1381 TTCCACTCCCAACGACATGTCTTCAGGACTACTTGTGGCAGCGGTCTCTCATCTAC 1440
 QY 481 AsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAlaGlyLeuLeuVal 500
 DB 1441 AACACGGGTTTCATGTGCTTGGCAGGACTTGGACCCCAACACCGCGGGTGTGTGTG 1500
 QY 501 AsnTrpProLysTyrThrSerSerSerGlnSerGlyAsnAsnLeuMetMetIleAsnAla 520
 DB 1501 AAGTGGCCCGAGTACACACGAGCTCTCAGTCTGGCAACAACCTTGATGATGATCAACGCC 1560
 QY 521 LeuGlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGlyTyrAspAlaLeuMetThr 540
 DB 1561 TTGGGCTTGTACACGGCAAGACAACTCCCGCACCGCGGTACGACGCTGTGTCTCC 1620
 QY 541 AsnProSerSerPhePheVal 547
 DB 1621 AACCCGCGCTCTTCTTTGTG 1641

RESULT 3
 ID AAX33112 standard; DNA; 1647 BP.

XX AAX33112;
 XX 23-JUN-1999 (first entry)

Synthetic lipase 1 gene.

Candida rugosa; lipase 1; LIP1; industrial bioconversion; ss.

Synthetic.
 Candida rugosa.

WO9914338-A1.

25-MAR-1999.

16-SEP-1997; 97WO-NL000524.

16-SEP-1997; 97WO-NL000524.

(UNIL) UNILEVER NV.

Brocca S, Schmidt-Dannert C, Lotti M, Alberghina L, Schmid R;

WPI; 1999-229539/19.

Synthesis and functional overexpression of a Candida rugosa lipase gene coding for a major industrial lipase.

Disclosure; Page 23-28; 44pp; English.

The present sequence represents a synthetic lipase 1 gene derived from the native Candida rugosa lipase 1 gene. Lipases produced by Candida rugosa are extensively used in industrial bioconversions, and the pure lipase 1 can be used in a process requiring high specificity toward acyl chains shorter than 14C. Lipase 1, free of 2-5, can be obtained without using expensive and expensive working up procedures. Pure lipase 1 exhibits higher activity toward capriate than toward palmitate

XX SQ Sequence 1647 BP; 377 A; 400 C; 364 G; 506 T; 0 U; 0 Other;
 Alignment Scores: 9.9e-230 Length: 1647
 Pred. No.: 2513.00 Matches: 470
 Score: 93.48% Conservative: 32
 Percent Similarity: 87.52% Mismatches: 35
 Best Local Similarity: 87.74% Indels: 0
 Query Match: 87.74% Gaps: 0
 DB: 2
 US-09-943-857-4 (1-547) x AAX33112 (1-1647)
 QY 11 LeuGlySerValProThrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAla 30
 DB 37 GTTCTGCTGCTGCCCAACCGGCACCTTGGCTAACGGTGACACCATCCGGGTTTGAAGGCC 96
 QY 31 IleIleAsnGluAlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArg 50
 DB 97 ATCATCAACGAAGCCTTCTTGGGTATTCCATTTCCGAACCAACAGTTGGTAACTTGAGA 156
 QY 51 PheLysAspProValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGly 70
 DB 157 TTCAGGACCCAGTTTCCATCTCCGTTCTTGGATGGTCAAAAGTTTCACTTCTTACGGT 216
 QY 71 ProSerCysMetGlnGlnAsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAla 90
 DB 217 CCATCTTGTATGCAACAAACCCAGAGGTACTACGAAGAAACTTCCCAAGACGCT 276
 QY 91 LeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCys 110
 DB 277 TTAGATCTGGTTATGCAATCCAAAGTTTTCGAAGCTGTTTCTCCATCTTCTGAAGACTGT 336
 QY 111 LeuThrIleAsnValValArgProGlyThrLysAlaGlyAlaAsnLeuProValMet 130
 DB 337 TTGACCATTAATGTTGTAGACCACCCGCGACAAAGGCTGGTGCCAACTTGCACGTATG 396
 QY 131 LeuTrpIlePheGlyGlyPheGluIleGlySerProThrIlePheProProAlaGln 150
 DB 397 TTGTGGATCTTGTGGTGGTGTGTTTGAAGTTGGTGTACTAGTACCTTCCCTCCAGCCCAA 456
 QY 151 MetValThrLysSerValLeuMetGlyLysHisIleIleHisValAlaValAsnTyrArg 170
 DB 457 ATGATTACCAAGTCTATTGCTATGGTAAGCAATCAATCCACGTTTCTGTCACTACAGA 516
 QY 171 ValAlaSerTrpGlyPheLeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsnAla 190
 DB 517 GTCCTGAGCTGGGGTTCCTTGGCTGGTACGAAATCAAGCCGGAAGTTCTGCCAACGCC 576
 QY 191 GlyLeuLysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGly 210
 DB 577 GGTTTGAAGGACCAAGATTGGGTATGCAATGGGTGGCTGACAACTTGTCTTGTGTGT 636
 QY 211 GlyAspProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCys 230
 DB 637 GGTGATCCCAACTAAGGTTACTATCTTTGGTGAATCTGCTGGTTCATGTCCTGCTCATGTGT 696
 QY 231 HisLeuIleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGly 250
 DB 697 CACATTTTGTGGAACGCGGTGACAACTTACAGGGTAAAGCAATGTTTACAGAGTGTGT 756
 QY 251 IleMetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGlu 270
 DB 757 ATCATGCAATCTGGTGTATGGTTCCATCTGACGCGGTGACGGTATCTACGGTAAACGAA 816
 QY 271 IleTyrAspLeuPheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCys 290
 DB 817 ATTTTTGACTTGTGGCTTCCACGCTGGTGTGGTGTCTGCTCTTCGCAAGTTGGCTGTGT 876
 QY 291 LeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeu 310
 DB 877 TTGAGAGGTGTTTCTTCTGACACTTGGAGAGCGCCACCAACACCCCTGGTTCCTTG 936

```

QY 311 AlaTyrSerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThrAsp 330
Db 937 GCTTACTCTCTCTTAGATTGTCTTACTTGCCAAAGACACGCGTGTAAACATCACCGAC 996
QY 331 AspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleIleGlyAsp 350
Db 997 GACATGTACGCTTGTGTAGAGAGGTAAAGTATGCCAACATCCCTGTATCATCGGTGAC 1056
QY 351 GlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnAlaGln 370
Db 1057 CAAACGACGAGGTACTCTTTGTGTACTTCTTCTTTGAAGCTTACCACGTATGCCAA 1116
QY 371 AlaArgAlaTyrPheLysGlnSerPheIleHleAlaSerAspAlaGluIleAspThrLeu 390
Db 1117 GCAGAGAAATATTCAAGCAATCTTTGTCCAGCTAGCGACGCTGAAATCGACACTTTG 1176
QY 391 MetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsn 410
Db 1177 ATGACTGCTTACCAGGTGACATCACTCAAGGTTCTCCATTTGACACTGGAATCTTAAC 1236
QY 411 AlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHis 430
Db 1237 GCCTTGACCCCAATTCAGAGAAATCTGCTGTTTGGGTGACTTGGGTTTACTTTG 1296
QY 431 AlaArgArgTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSerLys 450
Db 1297 GCTGCTAGATATCTTCTGAACCACTACACCGGTGTACCAAGTACTCTTTCTGTCTAAG 1356
QY 451 GlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAsp 470
Db 1357 CAATGCTGTGTTTGCAGTTTGGGTGACTTCCATCCCAACGATATCGCTTCCAAAGAC 1416
QY 471 TyrLeuLeuGlySerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAsp 490
Db 1417 TACTTGTGGTTCGTGTTCCCTTGAICTACAAACGCTTTCATTCGTTTGCCTGAC 1476
QY 491 LeuAspProAsnThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerGln 510
Db 1477 TTGGACCCAAACACCGCGGTTTGTGTTAAGTGGCCAGATACACCTCTTCTCTCAA 1536
QY 511 SerGlyAsnAsnLeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPhe 530
Db 1537 TCTGGTAAACACTTGAATGATCAACGCTTGTGGGTTGTACACCGGTGAAGCAACTTC 1596
QY 531 ArgThrAlaGlyTyrAspAlaLeuMetThrAsnProSerSerPhePheVal 547
Db 1597 AGAACCGCGGTTACGACGCTTGTCTCCAAACCCACCTCTTCTTGTGTT 1647

RESULT 4
AAH43625
ID AAH43625 standard; cDNA; 1950 BP.
XX
AC AAH43625;
XX
DT 21-JAN-2002 (first entry)
XX
DE Synthetic lip1 gene.
XX
KW Lipase; lip1; variant; ripening form; serine; 16-18C acyl chain; ss.
XX
OS Candida rugosa.
XX
FH Key Location/Qualifiers
FT CDS /*tag= a
FT /*product= "lip1"
FT /transl_except= (pos:292..294,aa:Glu)
FT /transl_except= (pos:301..303,aa:Gln)
FT /transl_except= (pos:532..534,aa:Thr)
FT /transl_except= (pos:553..555,aa:Thr)
FT /transl_except= (pos:565..567,aa:Arg)
FT /transl_except= (pos:802..807,aa:Thr)
FT /transl_except= (pos:1024..1026,aa:Glu)

```

```

FT /transl_except= (pos:1114..1116,aa:Thr)
FT /transl_except= (pos:1288..1290,aa:Thr)
FT /transl_except= (pos:1678..1680,aa:Gly)
FT /transl_except= (pos:1753..1755,aa:Phe)
XX
PN EP1130100-A1.
XX
PD 05-SEP-2001.
XX
PF 02-FEB-2001; 2001EP-00200375.
XX
PR 14-FEB-2000; 2000EP-00200513.
XX
PA (UNIL ) UNILEVER NV.
PA (UNIL ) UNILEVER PLC.
XX
PI Brocca S, Bornscheuer UT, Pleiss J, Schmid RD, Schmid U;
PI Schmitt J;
XX
DR WPI; 2001-649825/75.
DR P-PSDB; AAB47622.
XX
PT Modified lipolytic enzymes with altered substrate specificity, useful for
PT biocatalytic applications comprising high specificity towards carbon 16
PT and carbon 18 acyl chains.
XX
PS Disclosure; Fig 1; 33pp; English.
XX
CC This sequence encodes a parent lipase, lip1, derived from C. rugosa. The
CC lipase of the invention is a variant of this parent lipase, with altered
CC properties. The variant is the ripening form of C. rugosa lipase selected
CC from pre, pro, prepro or mature lipase, in which 60% or less of the CTG
CC codons encoding serine in the native C. rugosa sequence, are replaced by a
CC universal codon for serine. The modified nucleic acid sequence is further
CC modified, such that lipase variant exhibits an altered property. The
CC modified lipase is useful in a process requiring high specificity towards
CC 16-18C acyl chains
XX
SQ Sequence 1950 BP; 472 A; 478 C; 420 G; 580 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.55e-229 Length: 1950
Score: 2512.00 Matches: 470
Percent Similarity: 93.64% Conservative: 31
Best Local Similarity: 87.85% Mismatches: 34
Query Match: 87.71% Indels: 0
DB: Gaps: 4

US-09-943-857-4 (1-547) x AAH43625 (1-1950)
QY 13 SerValProThrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAlaIleIle 32
Db 172 GCTGCCCAACCGCCACCTTGTGCTAACGGTGACCATCCCGGTTGAACCGCATCATC 231
QY 33 AsnGluAlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArgPheLys 52
Db 232 AACGAAGCCTTCTGGGTATTCATTTCCGAACCAACCAAGTTGGTAACTTGAGATTCAAG 291
QY 53 AspProValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyProSer 72
Db 292 GACCCAGTTCCATACTCCGGTTCCTTGGATGTCGTAAGTTACATCTCTTACGGTCCATCT 351
QY 73 CysMetGlnGlnAsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAlaLeuAsp 92
Db 352 TGTATGCAACAAACCCAGAGGTACTACGAAGAAACTTCCCAAGAGGACGCTTTAGAT 411
QY 93 LeuValMetGlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCysLeuThr 112
Db 412 CTGGTTATGCAATCCAAAGTTTTCGAAGCTGTTTCTCCATCTTCTCGAAGACGTTTGACC 471
QY 113 IleAsnValValArgProGlyThrLysAlaGlyAlaAsnLeuProValMetLeuTrp 132
Db 472 ATTAATGTTGTAGACACCGGGGCAAAAGGCTGGTGGCAACTTCCAGTTATGTTGTGG 531

```


Query Match:	87.12%	Indels:	0
DB:	7	Gaps:	0
US-09-943-857-4 (1-547) x ABX95908 (1-1641)			
QY	1	SerMetAsnSerArgGlyProAlaGlyArgLeuGlySerValProThrAlaLysLeuAla	20
DB	1	TCATGAATTCAGGTGGCCAGGCGCGGTCTCGGATCGGTACCCCACTGCCAGCGTCGCC	60
QY	21	AsnGlyAspThrIleThrGlyLeuAsnAlaIleleAsnGluAlaPheLeuGlyIlePro	40
DB	61	ACGGCGACACCATCACCGGTCTCAAGCCCATCATCAACGAGGCGTTCCTCGGCATTCCC	120
QY	41	PheAlaGluProProValGlyAsnLeuArgPheLysAspProValProTyrSerGlySer	60
DB	121	TTTGCCGAGCGCGGTGGCAACCTCCGGTTCAGGACCCCTGTGCGGTACCGTGGGTCT	180
QY	61	LeuAsnGlyGlnLysPheThrSerTyrGlyProSerCysMetGlnGlnAsnProGluGly	80
DB	181	CTCAACGGTCAATCTCTCAACCGGTACGGTCCGTCTTGCATGACGACAGAACCCCGAGGC	240
QY	81	ThrPheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPhe	100
DB	241	ACCTACGAGGAGACCTCCCAAGGTGGCGCTTGACTTGGTGATGCAGTCCAAAGGTGTC	300
QY	101	GlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgProProGly	120
DB	301	CAGGCTGTCTCCCAACAGCGAGGACTGCCTCACCATCAACGTGGTGGCGCGCGGC	360
QY	121	ThrLysAlaGlyAlaAsnLeuProValMetLeuThrIlePheGlyGlyPheGluIle	140
DB	361	ACCAAGCGCGCGCCAACTCCCGGTTCATGCTCTGGATCTTTGGCGGTGGGTTTGATC	420
QY	141	GlySerProThrIlePheProAlaGlnMetValThrLysSerValLeuMetGlyLys	160
DB	421	GGCAGCCCCACCATCTTCCCTCCCGCTCAGATGGTCTCCNAGAGTGTGCTCATGGCGAG	480
QY	161	HisIleIleHisValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGlyAsp	180
DB	481	CCCATCATCAGTGGCGCTCAACTACCGCTTGGCGTCTTTGGTTCTTGGCGGTCCG	540
QY	181	AspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMetGln	200
DB	541	GACATCAAGCGCGAGGCGAGCTCCATGGCGGCTCAAGACCAGCGCTTGGCGATGCAG	600
QY	201	TrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysValThrIlePheGly	220
DB	601	TGGGTGGCGAGACNACATTCCCGGTTCCGGCGCACCCGAGCAAGGTGACCATCTTGGC	660
QY	221	GluSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAspGlyAspAsnThr	240
DB	661	GAGTCTGCGCGGAGCATGTCCGTGTTGTGTCACCTTCTTGGATGGCGGCGACACACG	720
QY	241	TyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValProSer	260
DB	721	TACAAGGCAAGCCGTGTGTCCGCGCGGCGATCATGAGTCTGGAGCCATGTGGCGTCT	780
QY	261	AspProValAspGlyThrTyrGlyAsnGluIleTyrAspLeuPheValSerSerAlaGly	280
DB	781	GACCCGTGGACGGACCTATGGAGCCCAATCATGACACGTGTGGTGGCTTCTACGGGC	840
QY	281	CysGlySerAlaSerAspLysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeu	300
DB	841	TGCAGCAGTGCAGCAACAAGCTTGGCTTGGTGGTCTTCTTACTCAGGCATTGCTC	900
QY	301	AspAlaThrAsnAsnThrProGlyPheLeuAlaTyrSerSerLeuArgLeuSerTyrLeu	320
DB	901	GATGCCACCAACACACCCCTGGTCTTGTGTGTATACACCTCGTGGGTGTTCTTATCTC	960
QY	321	ProArgProAspGlyLysAsnIleThrAspMetTyrLysLeuValArgAspGlyLys	340
DB	961	CGCGGCGCGAGCGGCCCAACATCACCGATGACATGTACAGTTGTGCGGACGCAAG	1020
QY	341	TyrAlaSerValProValIleIleGlyAspGlnAsnAspGluGlyThrIlePheGlyLeu	360
DB	1021	TATGCAAGCGTTCCTCGGTGATCATTTGGGACCAACGACGAGGCTCTCTTTGTGATCTC	1080
QY	361	SerSerLeuAsnValThrThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPheIle	380
DB	1081	TCCTTCTTTGAACACCAACCGAGGCGGACCGCGAGCATACCTCAGAAAGTCTTTTCATC	1140
QY	381	HisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyrProGlnAspIleThrGln	400
DB	1141	CACGCCACCGACCGCATATCACCGCATTAAGGCGGCTACCCAGCGATGTACCCAG	1200
QY	401	GlySerProPheAspThrGlyValLeuAsnAlaLeuThrProGlnPheLysArgIleSer	420
DB	1201	GGTTCCTCGGTTCGACACGGGCATCTCAAGCCCTTACACCCGCTCAAGCGGATCAAT	1260
QY	421	AlaValLeuGlyAspLeuAlaPheIleHisAlaArgArgTyrPheLeuAsnHisPheGln	440
DB	1261	GCTGTGCTTGGCGACCTTACCTTACCTCTCTAAGACCTTTCTGGGTGGCCATTTCTGTG	1320
QY	441	GlyGlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeuProIleMetGlyThr	460
DB	1321	GGTGTGCTTGGCGACCTTACCTTACCTCTCTAAGACCTTTCTGGGTGGCCATTTCTGTG	1380
QY	461	PheHisAlaAsnAspIleValTrpGlnAspTyrLeuLeuGlySerGlySerValIleTyr	480
DB	1381	TTCCACGCGAAGACATTTGTGGCAGCACTTTTGTGGCAGCGCGAGCGTCACTATC	1440
QY	481	AsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAlaGlyLeuLeuVal	500
DB	1441	AACAACGCGTTCATCGCGTTTGCACCGCACTTGGACCCCAACACCGCGGCTTGTCTGTG	1500
QY	501	AsnTrpProLysTyrThrSerSerSerGlnSerGlyAsnAsnLeuMetMetIleAsnAla	520
DB	1501	CAGTGGCCCCAGTACACAGCAGCTCTCAGCGGGGAGCAACTTGTATGAGATCAGTGCC	1560
QY	521	LeuGlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGlyTyrAspAlaLeuMetThr	540
DB	1561	TTGGGCTTGTACACCGGCAAGCAACATTCGCGACCGCGGCTACCAACGCTTGTGTGCC	1620
QY	541	AsnProSerSerPhePheVal	547
DB	1621	GACCCGCTCTCACTTTTTCGTG	1641
RESULT 6			
ID	AA33111	standard; DNA; 1649 BP.	
XX	AA33111		
AC	AA33111		
DT	23-JUN-1999	(first entry)	
XX			
DE		Candida rugosa lipase 1 gene.	
XX			
XX		Candida rugosa; lipase 1; LIP1; industrial bioconversion; ss.	
OS		Candida rugosa.	
XX			
PN	WO9914338-A1.		
XX			
PD		25-MAR-1999.	
XX			
PF		16-SEP-1997; 97WO-NL000524.	
XX			
PR		16-SEP-1997; 97WO-NL000524.	
XX			
PA		(UNIL) UNILEVER NV.	
XX			
PI		Brocca S, Schmidt-Dannert C, Lotti M, Alberghina L, Schmid R;	
XX		WPI; 1999-229539/19.	
XX			
PT		Synthesis and functional overexpression of a Candida rugosa lipase gene	


```

FH Key Location/Qualifiers
FT CDS 1..1641
FT FT /*tag= a
FT FT /product= "Lipase 4"
FT FT /partial
FT FT /note= "No start or stop codon shown"
PN BP1288294-A2.
XX XX
XX 05-MAR-2003.
XX XX
XX 26-APR-2002; 2002EP-00009616.
XX XX
XX 31-AUG-2001; 2001US-00943857.
XX XX
XX (SINI-) ACAD SINICA.
XX XX
XX Tang S, Lee G, Shaw J;
XX XX
XX WPI; 2003-395476/38.
XX XX
XX P-PSDB; ABU09072.
XX XX
XX Isolated mutant nucleic acid encoding Candida rugosa lipase, useful for
XX the preparation of Candida rugosa lipase for biocatalytic applications.
XX Disclosure; Page 7-8; 33pp; English.
XX XX
XX The invention relates to an isolated mutant nucleic acid encoding a
XX Candida rugosa lipase polypeptide. The DNA has a sequence having at least
XX 80% identity to a wild-type DNA encoding Candida rugosa lipase. The
XX invention also relates to a microorganism comprising the DNA, where the
XX microorganism is a bacterium or yeast, preparing a mutant DNA encoding a
XX C. rugosa lipase and a chimeric C. rugosa lipase comprising a substrate
XX interacting domain of a first C. rugosa lipase and a non-substrate
XX interacting domain of a second C. rugosa lipase. The method is useful for
XX preparing a mutant DNA encoding a Candida rugosa lipase. The DNA is
XX useful in the large scale manufacture of Candida rugosa lipase which is
XX useful for biocatalytic applications. This sequence represents DNA
XX encoding Candida rugosa lipase 4
XX XX
XX Sequence 1641 BP; 306 A; 524 C; 461 G; 350 T; 0 U; 0 Other;
XX XX
XX Alignment Scores:
XX Pred. No.: 3..39e-222 Length: 1641
XX Score: 2434.00 Matches: 457
XX Percent Similarity: 90.49% Conservative: 38
XX Best Local Similarity: 83.55% Mismatches: 52
XX Query Match: 84.99% Indels: 0
XX DB: 7 Gaps: 0
XX
XX US-09-943-857-4 (1-547) x ABX95907 (1-1641)
XX
XX QY 1 SerMetAsnSerArgGlyProAlaGlyArgLeuGlySerValProThrAlaLysLeuAla 20
XX Db 1 TCATGAATTCACGTGCGCCAGCGCGCGTCTCGATCGGTACCCACTGCCACGCTCGCC 60
XX
XX QY 21 AsnGlyAspThrIleThrGlyLeuAsnAlaIleAsnGluAlaPheLeuGlyIlePro 40
XX Db 61 AACGGGACACCATCACCGGTCTCACGCCATCATCAACAGGCGGTCTCGGTATTC 120
XX
XX QY 41 PheAlaGluProProValGlyAsnLeuArgPheLysAspProValProTyrSerGlySer 60
XX Db 121 TTTCGTCAGCGCGCGGTGGGCAACCTCCGCTTCAAGCGGCTGTGCGGTACTCGGCGTCT 180
XX
XX QY 61 LeuAsnGlnLysPheThrSerTyrGlyProSerCysMetGlnGlnAsnProGluGly 80
XX Db 181 CTCAATGGTCAGAAGTTTACTTCGTATGCGCCCTTCGTGATGCAGATGAACCCATTGGGC 240
XX
XX QY 81 ThrPheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPhe 100
XX Db 241 AACTGGGACTCTCGCTTCCCAAGGCTGCCATCACTCTTGAAGCAGTCCCAAGCTCTTC 300
XX
XX QY 101 GlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgProGly 120

```

```

Db 301 CAGCGGTGCTTCTTAACGGCGAGGACTGTCCTACCATCAACGTTGGTGGCGCGTCAGGC 360
QY 121 ThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyGlyGlyPheGluIle 140
Db 361 ACCAAGCGGGTGCCAAACCTCCCGTGATGTGTGGATTTTGGCGCGGGTTTCAGGTT 420
QY 141 GlySerProThrIlePheProAlaGlnMetValThrLysSerValLeuMetGlyLys 160
Db 421 GCGGCTCCAGTCCTCTCCCTCCGACACAGATGATACCCGCGCGGTCTTATGGCAAG 480
QY 161 HisIleIleHisValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGlyAsp 180
Db 481 CCCATCATCCACGTGACGATGAACCTACCGGTTGCTTCGTGGGGTCTTGGCTGGTCCA 540
QY 181 AspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMetGln 200
Db 541 GACATCAAGGCCGAGGCGACGCGGACCGCGGTTTGACACCAACGCTTGGGTTTCAG 600
QY 201 TrpValAlaAspAsnIleAlaGlyPheGlyAspProSerLysValThrIlePheGly 220
Db 601 TGGGTGGCGGACCAACATTGCGGGTTCGGCGGACCCGTCGAAGGTGACCACTTTGGT 660
QY 221 GluSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAspGlyAspAsnThr 240
Db 661 GAGTCGGCGGCGACGATGTCGGTAATGTGTGCTGTCAGCTCTCTGGACGCGCGCAACACG 720
QY 241 TyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValProSer 260
Db 721 TACAACGGCAAGCGGTTGTTCCGTGCGCCATCATGCAGTCTGGGGCATGGTGGCTCG 780
QY 261 AspProValAspGlyThrTyrGlyAsnGluIleTyrAspLeuPheValSerSerAlaGly 280
Db 781 GACCCGGTGGATGGCCCTACGCGACGAGATCTACACCGAGTGTTCGTCAGCGCGG 840
QY 281 CysGlySerAlaSerAspLysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeu 300
Db 841 TGTGGCAGTGCACGACGACAGCTCGGTGCTTCGCGCAGCATCTCGAAGCAACACTTTC 900
QY 301 AspAlaThrAsnAsnThrProGlyPheLeuAlaTyrSerSerLeuArgLeuSerTyrLeu 320
Db 901 CAGGCCACCGACGACACTCCGGGGGCTTGGCGTACCCCTCGTTGGCGTTGTGCTTCTC 960
QY 321 ProArgProAspGlyLysAsnIleThrAspAspMetTyrLysLeuValArgAspGlyLys 340
Db 961 CCGCGCGCGCGACCGACCTTCATCCAGGATGATGTTCAAGTGTGGTGGCGCGGCAAG 1020
QY 341 TyrAlaSerValProValIleIleGlyAspGlnAsnAspGluGlyThrIlePheGlyLeu 360
Db 1021 TGTGCCAACGTTCCGGTGATCATTTGGCGACCAAGACGAGGCGGACAGTGTTCGGTTG 1080
QY 361 SerSerLeuAsnValThrThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPheIle 380
Db 1081 TCCAGTTTGAACGTGACTACGATGCTCAGGCAAGCCAGTACTTCAAGGAAGCTTCATC 1140
QY 381 HisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyrProGlnAspIleThrGln 400
Db 1141 CAGCCGAGGACCGCGAGATCGACCTTGATGGCGGCTACCCCGGCGACATCACCCAG 1200
QY 401 GlySerProPheAspThrGlyValLeuAsnAlaLeuThrProGlnPheLysArgIleSer 420
Db 1201 GGTAGTCCGTCGACACCGGCTCTTCAAGCCCATCATCCCGCGAGTTCAAACGGATGCA 1260
QY 421 AlaValLeuGlyAspLeuAlaPheIleHisAlaArgArgTyrPheLeuAsnHisPheGln 440
Db 1261 GCGGTGCTTGGTGACCTTCGCTCCCGCGCGCTACTTCTCAACACACTTCCAG 1320
QY 441 GlyGlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeuProIleMetGlyThr 460
Db 1321 GCGGGACCAAGTACTCGTTCCTCTGAAGAGCTTAGTGGGTGCGGTGATGGCACCC 1380
QY 461 PheHisAlaAsnAspIleValTrpGlnAspTyrLeuLeuGlySerGlySerValIleTyr 480

```



```

QY 331 AspMetTyrIysLeuValArgAspGlyLysTyrAlaSerValProValIleIleGlyAsp 350
Db 997 GACATGTACCGCTTGTGGCGCAGGCAAGTATGCCAACATCCCTGTGATCATCGCGCAC 1056
QY 351 GluAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrAsnAlaGln 370
Db 1057 CAGAACACGAGGAGCCTTCTTTGGCACCCCTGCTGTGAACGTGACCGGATGCCCGAG 1116
QY 371 AlaArgAlaTyrPheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeu 390
Db 1117 GCCCGGAGTACTTCAAGCAGCTGTGTTCACGCCAGCGAGCGGAGATCGACACGTTG 1176
QY 391 MetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsn 410
Db 1177 ATGACGGCGTACCCCGCGCATACCCAGCGCGCTGCCGTTCGACACGGGTATCTCAAC 1236
QY 411 AlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHis 430
Db 1237 GCCCTCACCCCGAGTTCAAGAGAACTCTGGCGGTGCTCGCGGACCTTGGCTTTACGCTT 1296
QY 431 AlaArgArgTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSerLys 450
Db 1297 GCTCGTGCCTACTTCTCAACCACTACACCGCGGCGACCAAGTACTCTATTCCTCCTGAAG 1356
QY 451 GlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAsp 470
Db 1357 CAGCTCTCGGCTTGGCGGTGCTCGGAACGTTCCACTCCCAACGACATTGTCTTCCAGGAC 1416
QY 471 TyrLeuLeuGlySerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAsp 490
Db 1417 TACTTGTGGGACGCGCTCGCTCATCTACAACAACGCTCATTTGGCTTTGCCACGGAC 1476
QY 491 LeuAspProAsnThrAlaGlyLeuLeuValAsnTrpProLysTyrTrpSerSerGln 510
Db 1477 TTGGACCCCAACACCGCGGGTGTGTGTGAAGTGGCCCGAGTACACCACGACGCTGCAG 1536
QY 511 SerGlyAsnAsnLeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPhe 530
Db 1537 CTGGGCAACAACCTGTATGATGATCAACGCTTGGGCTTGTACACCGCGCAAGGACAACTTC 1596
QY 531 ArgThrAlaGlyTyrAspAlaLeuMetThrAsnProSerSerPhePheVal 547
Db 1597 CGCACCGCGGCTACGACGCTTGTCTTCCAAACCGCGCGCTGTCTTGTG 1647

RESULT 9
ID ABX95905
XX ABX95905 standard; DNA; 1641 BP.
AC ABX95905;
XX
XX 15-JUL-2003 (first entry)
XX
DE Candida rugosa lipase 2 DNA.
XX
XX Lipase 2; gene; ds; mutant.
XX
OS Candida rugosa.
XX
FH Key
FT Location/Qualifiers
FT 1. .1641
FT /*tag= a
FT /product= "Lipase 2"
FT /partial
FT /note= "No start or stop codon shown"
XX
PN EP1288294-A2.
XX
XX 05-MAR-2003.
PD
XX
XX 26-APR-2002; 2002BP-00009616.
PF
XX
XX 31-AUG-2001; 2001US-00943857.
PR
XX

```

(SINI-) ACAD SINICA.

Tang S, Lee G, Shaw J;

WPI; 2003-395476/38.

P-PSDB; ABU09070.

Isolated mutant nucleic acid encoding Candida rugosa lipase, useful for the preparation of Candida rugosa lipase for biocatalytic applications.

Claim 25; Page 4-5; 33pp; English.

The invention relates to an isolated mutant nucleic acid encoding a Candida rugosa lipase polypeptide. The DNA has a sequence having at least 80% identity to a wild-type DNA encoding Candida rugosa lipase. The invention also relates to a microorganism comprising the DNA, where the microorganism is a bacterium or yeast, preparing a mutant DNA encoding a C. rugosa lipase and a chimeric C. rugosa lipase comprising a substrate interacting domain of a first C. rugosa lipase and a non-substrate interacting domain of a second C. rugosa lipase. The method is useful for preparing a mutant DNA encoding a Candida rugosa lipase. The DNA is useful in the large scale manufacture of Candida rugosa lipase which is useful for biocatalytic applications. This sequence represents DNA encoding Candida rugosa lipase 2

Sequence 1641 BP; 305 A; 521 C; 472 G; 343 T; 0 U; 0 Other;

Alignment Scores:

```

Pred. No.:      1,43e-218      Length:      1641
Score:          2396.00         Matches:    450
Percent Similarity: 88.48%      Conservative: 34
Best Local Similarity: 82.27%   Mismatches:  63
Query Match:     83.66%        Indels:     0
DB:              7            Gaps:         0

```

US-09-943-857-4 (1-547) x ABX95905 (1-1641)

QY 1 SerMetAsnSerArgGlyProAlaGlyArgLeuGlySerValProThrAlaLysLeuAla 20

Db 1 TCATGAATTCACGTGGCCCGCGCGCTCGGATCGGTATCCACCGCCGCTCGCC 60

QY 21 AsnGlyAspThrIleThrGlyLeuAsnAlaIleIleAsnGluAlaPheLeuGlyIlePro 40

Db 61 AACGGCGACACCATCACCGGTCTCAACGCCATTGTCAAGAAAGTTCTCGCATACCG 120

QY 41 PheAlaGluProProValGlyAsnLeuArgPheLysAspProValProTyrSerGlySer 60

Db 121 TTGGCGAGCGCGCGCTGGGCGACCTCCGCTTCAAGCGCGCTACTCGGCGTCG 180

QY 61 LeuAsnGlyGlnLysPheThrSerTyrGlyProSerCysMetGlnGlnAsnProGluGly 80

Db 181 CTCACGGCGCAGGAGTTTACCTCTTACGGCCCGCTTGTATGATGATGAACCTATGGGC 240

QY 81 ThrPheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPhe 100

Db 241 TCGTTTGAGGACACACTTCCCAAGATGCGTTGACTTGGTGTCTCCAGTCCAAGATCTTC 300

QY 101 GlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgProGly 120

Db 301 CAAGTGTGTCTTCCCAACGACGAGGACTGTCTCACCATCAAGTGTATCGGCGCGCGC 360

QY 121 ThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyClyGlyPheGluIle 140

Db 361 ACCAGGCCAGTGTCTGCTTCCCGGTGATGTCTTGGATCTTTGGCGGTGGGTGTGAGCTT 420

QY 141 GlySerProThrIlePheProProAlaGlnMetValThrLysSerValLeuMetGlyLys 160

Db 421 GGCGGCTCCAGCCTCTTTCAGGAGACACAGATGGTGGCCAAAGAGCGTGTCTATGGTAA 480

QY 161 HisIleIleHisValAlaValAsnTyrArgValAlaSerTyrGlyPheLeuAlaGlyAsp 180

Db 481 CCGGTGATCCAGCTGAGCATGAATACCGGTGGGTGATCGGGGTCTTGTGGCGGCC 540

181 AspIleLysAlaGlnGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMetGln 200
|||||
541 GACATCCAGAACGAGGAGCGGAGACCCCGCTTGATGACCGAGCGGTGGCCATCGAG 600
201 TrpValAlaAspAsnIleAlaGlyPheGlyAspProSerLysValThrIlePheGly 220
|||||
601 TGGGTGGCGGACACATTTGCTGGGTGGCGCGACCGACGACGAGGTGACCATATACGC 660
221 GluSerAlaGlySerMetSerValLeuCysHisIleuLeuTrpAsnAspGlyAspAsnThr 240
|||||
661 GAGTCTGGCGGACGATGTCGACGTTTGTGACCTTGTGTGAAACGACGCGGACACACG 720
241 TyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValProSer 260
|||||
721 TACAACGGCAGACCGCTTTCGCGCCGCCATCATGCGTGGTGTGATGGTGGCTCT 780
261 AspProValAspGlyThrTyrGlyAsnGluIleTyrAspLeuPheValSerSerAlaGly 280
|||||
781 GACCGCGTGGACGCGACGTCAGCGACCGGAGATCTACAACCGAGGTGCTGGCGTCTGCCGG 840
281 CysGlySerAlaSerAspLysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeu 300
|||||
841 TGTGCACTGCGCGACGACGATCGCTGCTTTCGCGCGCTTCTCAGGACACGTTGTAC 900
301 AspAlaThrAsnAsnThrProGlyPheLeuAlaTyrSerSerLeuArgLeuSerTyrLeu 320
|||||
901 CAGGCCACGAGCGACCGCCGCGGTGGGTATACCGTCTGTCGGGTGCTTATCTC 960
321 ProArgProAspGlyLysAsnIleThrAspMetTyrLysLeuValArgAspGlyLys 340
|||||
961 CGCGCGCCGACGCGACCTTCTACCGGACGACATGTATGCCCTTGTGCGGACGCGCAG 1020
341 TyrAlaSerValProValIleIleGlyAspGlnAsnAspGluGlyThrIlePheGlyLeu 360
|||||
1021 TACGCACACGTCGCGGTGATCATCGCGACCGACGAGCGAGCGACTTGTGGGCTC 1080
361 SerSerLeuAsnValThrThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPheIle 380
|||||
1081 TCTTTCTTTGACGTGACACACATGCTCAGGACCGGCGTACTTCAAGCAGCTTTTCATC 1140
381 HisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyrProGlnAspIleThrGln 400
|||||
1141 CACGCGACGATGCGGAGATCGACAGTTGATGGCGCGGTACACCGCGACATCACCCAG 1200
401 GlySerProPheAspThrGlyValLeuAsnAlaLeuThrProGlnPheLysArgIleSer 420
|||||
1201 GGTTCCTCCGTCGACACCGGATCTTCAATGGCATCCACCCGAGTTCAACGAGATCTCT 1260
421 AlaValLeuGlyAspLeuAlaPheIleHisAlaArgArgTyrPheLeuAsnHisPheGln 440
|||||
1261 GGTTCCTGGGACGCTTGGTTCACGCTTGGCGTGGCTGCTTCTCACTCACTTACCTAC 1320
441 GlyGlyThrIysThrSerPheLeuSerLysGlnLeuSerGlyLeuProIleMetGlyThr 460
|||||
1321 GCGCGCACCAAGTACTCGTCTCTCTAAGCAGCTTCTGGGTGGCGTGGCTTCTGGCACC 1380
461 PheHisAlaAsnAspIleValTrpGlnAspTyrLeuLeuGlySerGlySerValIleTyr 480
|||||
1381 TTCACGCGCAACACATCTCTGGCAGGACTCTTGGTGGCGAGCGGAGTGTGATCTAC 1440
481 AsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAlaGlyLeuLeuVal 500
|||||
1441 AACACGCGTTCATTGGTGTGCGACGCTTGCACCGCTTGCACCGGCGGCTTGTGACC 1500
501 AsnTrpProLysTyrThrSerSerSerGlnSerGlyAsnAsnLeuMetMetIleAsnAla 520
|||||
1501 AACTGGCCACGATACACGACGCTCTCAGTCTGGCAACAACCTTGATGACGATCAACGCG 1560
521 LeuGlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGlyTyrAspAlaLeuMetThr 540
|||||
1561 TTGGGGTTGTACACCGGCAAGGCAACTTCCGCGGATGCGTACAGCGCCCTCTTTTC 1620
541 AsnProSerSerPhePheVal 547

||||| 1621 AACCCGCCGCTCTTCTTTGTG 1641
RESULT 10
AAQ54020
ID AAQ54020 standard; DNA; 1828 BP.
XX AC AAQ54020;
XX 16-OCT-2003 (revised)
DT 25-MAR-2003 (revised)
DT 19-JUL-1994 (first entry)
XX Lipase coding sequence of Geotrichum candidum.
XX Immobilisation; enzyme; cell wall; alpha agglutinin; AGA 1; FLO 1;
KW Major cell wall protein; glycosyl-phosphatidyl-inositol;
KW anchoring protein; alpha factor; alpha-agglutinin; invertase; inulinase;
KW alpha-amylase; Geotrichum candidum; lipase; enzymatic process;
KW fermentation; biodegradation; catalysis; ss.
XX Galactomyces geotrichum.
FH Key Location/Qualifiers
FT CDS 40..1731
FT /tag= a
FT /product= "Lipase."
FT sig_peptide 40..96
FT /tag= b
FT mat_peptide 97..1728
FT /tag= c
FT /product= "Lipase."
XX WO9401567-A1.
PN 20-JAN-1994.
XX 07-JUL-1993; 93WO-EP001763.
XX 08-JUL-1992; 92EP-00202080.
PR 14-DEC-1992; 92EP-00203899.
XX (UNIL) UNILEVER PLC.
PA (UNIL) UNILEVER NV.
XX Klis FM, Schreuder MP, Toschka H, Verrips CT;
PI WPI; 1994-035071/04.
XX P-PSDB; AAR47577.
DR Immobilisation of enzymes to microbial cell wall - by prodn. of fusion
PT protein of enzyme linked to anchoring protein.
XX Claim 8; Page 49-52; 99pp; English.
CC The lipase is used in a method to immobilise enzymes to a microbial cell
CC wall. The coding sequence is used in the production of a recombinant
CC polynucleotide which comprises a structural gene encoding a protein with
CC catalytic activity (the lipase) and at least part of a gene encoding at
CC least the C-terminus of a protein capable of anchoring in a eukaryotic or
CC prokaryotic cell wall. The anchoring fragment or protein is selected from
CC alpha agglutinin, AGA 1, FLO 1, major cell wall protein of lower
CC eukaryotes or a proteinase of lactic acid bacteria. The recombinant
CC polynucleotide preferably also comprises a sequence encoding a signal
CC peptide to ensure secretion of the expressed product. The signal peptide
CC is preferably derived from glycosyl-phosphatidyl-inositol, anchoring
CC protein, alpha factor, alpha-agglutinin, invertase or inulinase, alpha-
CC amylase of Bacillus or proteinases of lactic acid bacteria. The host
CC microorganism can be used for performing enzymatic processes on an
CC industrial scale. (Updated on 25-MAR-2003 to correct PN field.) (Updated
CC on 16-OCT-2003 to standardise OS field)
XX Sequence 1828 BP; 389 A; 483 C; 419 G; 537 T; 0 U; 0 Other;
SQ

Alignment Scores:

Pred. No.: 1.56e-95 Length: 1828
 Score: 1107.00 Matches: 240
 Percent Similarity: 56.60% Conservative: 73
 Best Local Similarity: 43.40% Mismatches: 196
 Query Match: 38.65% Indels: 44
 DB: 2 Gaps: 9

US-09-943-857-4 (1-547) x AAQ54020 (1-1828)

QY 9 GlyArgLeuGlySerValProThrAlaLysLeuAlaLeuGlyAspThrIleThrGlyLeu 28
 DB 85 GGCACCTTGGCCAGCCCGCCAGCGCGTTCTTAATGGCAACAGAGTCATCTCTGGTGTG 144
 QY 29 AsnAlaIleAlaLeuGlyLeuGlyLeuProPheAlaGluProValGlyAsn 48
 DB 145 CTTGAGGGCAAGTTGATACCTTCAAGGAATCCATTGCTGACCTCTCTGTTGGTGAC 204
 QY 49 LeuArgPheLysAspProValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSer 68
 DB 205 TTGGCGTTCAAGCACCCCGCCAGCTTTCACCTGGATCTTACAGAGGCTCTTAAGGCCAACGAC 264
 QY 69 TyrGlyProSerCysMetGlnGlnAsnProGluGlyThrPheGlu 83
 DB 265 TTCAGCTCTGCTGTATGCGAGCTTGATCCTGGCAATGCCTTTCTTGTGTGACAAAGTA 324
 QY 84 -----GluAsnLeuGlyLysThrAlaLeuAspLeuValMet 95
 DB 325 GTGGGCTTGGAAAGATTCTTCGTGATACCTTAGAGGCCCTCTTATGACATGGCG--- 381
 QY 96 GlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsnVal 115
 DB 382 -----CAGGGTAGTGTCTCCATGAATGAGGACTGTCTCTACCTTAAGTT 426
 QY 116 ValArgProGlyThrLysAlaGlyAlaAsnLeuProValMetLeuThrIlePheGly 135
 DB 427 TTCGGCCCGCGTGGCCACCAAGCCGTGATGCTAAGCTCCCGCTCATGTTTGATTTACGGT 486
 QY 136 GlyGlyPheGluIleGlySerProThrIlePheProAlaGlnMetValThrLysSer 155
 DB 487 GGTGCTTGTGTGTTGTTCTTCTGCTTACCTCGTAAACGGCTACGTCAGAGGAGT 546
 QY 156 ValLeuMetGlyLysHisIleLeuHisValAlaValAsnTyrArgValAlaSerTrpGly 175
 DB 547 GTGAAATGGCCAGCGCTGTGTGTTGTTTCCATCACTACCGTACCGGCCCTATGGA 606
 QY 176 PheLeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGln 195
 DB 607 TTTCTGGGTGGTATGCCATCCATCCGCTGAGGGCAACCAACACGCTGCTGTCACGACCAG 666
 QY 196 ArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLys 215
 DB 667 CGCAAGGCTCGAGTGGGTGATGCGCAACATTGCCAATTTTGGTGGTATCCGCAAG 726
 QY 216 ValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsn 235
 DB 727 GTCATGATTTGGTGAGTCCGCTGGTGGCATGAGTGTGCTCACAGCTTGTGTGCTAC 786
 QY 236 AspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGly 255
 DB 787 GGTGGTGACAAACACCTACACGGAAGACGCTTTCCACTCTGCCATTCTTCAGTCTGGC 846
 QY 256 AlaMetValProSerAspProValAspGlyThrTyrGlyAsnGluIleThrAspLeuPhe 275
 DB 847 GGTCTCTTCTTACTTACTTACTTCTGTTGGTCCGAGAGTGCCTACAGCATTT 906
 QY 276 ValSerSerAlaGlyCys-----GlySerAlaSerAspLysLeuAlaCysLeuArg 292
 DB 907 GCTCAGTATCCGGATGTGACACCGACGAGTGCAGTATGACACTCTGGCTGTCTCCGC 966
 QY 293 SerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsn----- 305

Db 967 AGCAAGTCCAGCGATGCTTTGCACAGTGCAGAACTCGTATGATCTTAAGGACCTGTTT 1026
 QY 306 -----ThrProGlyPheLeuAlaTyrSerSerLeuArgLeuSerTyrLeuProArgPro 323
 Db 1027 GGTCTCTCTCCTCAATTCCTTGGATTGGT-----CCAGAGCCC 1065
 QY 324 AspGlyLysAsnIleThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSer 343
 Db 1066 GACGGCAACATTATTCGGATGCCGCTTATGAGCTCTACCGCAGCGGTAGATCGCAAG 1125
 QY 344 ValProValIleLeGlyAspGlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeu 363
 Db 1126 GTTCCCTTACATTACTGCAACACGAGGATGAGGTACTATTCTTGCCTCGTGTCTATT 1185
 QY 364 AsnValThrThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPheIleHisAlaSer 383
 Db 1186 AATGCTACCACTACTCCCATGTTAAGAGTGTGAAGTACATTGTAGCCAGGCTTCT 1245
 QY 384 AspAlaGluIleAspThrLeuMetAlaAlaTyrProGlnAspIleThrGlnGlySerPro 403
 Db 1246 GACGCTTCGCTTGTATCGTGTTCGCTCTACCCCGCTCTTGGTCGGAGGGTTCACCA 1305
 QY 404 PheAspThrGlyValLeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeu 423
 Db 1306 TTCGGCACTGGTATCTTAATGCTCTTACCCCTCAGTTCAAGCGCATTTGCTGCCATTTTC 1365
 QY 424 GlyAspLeuAlaPheIleHisAlaArgArgTyrPheLeuAsnHisPheGlnGlyGlyThr 443
 Db 1366 ACTGATTTGCTGTTCAGTCTCTCTCGTGTGTTATGCTTAACGCTACCAAGGACGTCAC 1425
 QY 444 LysTyrSerPheLeuSerLysGlnLeuSerGlyLeu---ProIleMetGlyThrPheHis 462
 Db 1426 CGTGGCATTTACCTTGGCCACCAAGCTCCATAACCTGTTCCATTTTGGGTACTTTCCAT 1485
 QY 463 AlaAsnAspIleValTrpGlnAspTyrLeu---LeuGlySerGlySerValIleTyrAsn 481
 Db 1486 GGCAGTGAATCTCTTTTCAATACTACTAGTGGACCTTGGCCCATCTTCTGTCT---TACCGC 1542
 QY 482 AsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAlaGlyLeuLeuValAsn 501
 Db 1543 CGTACTTTATCTCTGTTTGGCAACCAACGACGACCCCAACGTTTGTGTACCACTTCCACAG 1602
 QY 502 TrpProLysTyrThrSerSerGlnSerGlyAsnAsnLeuMetMetIleAsnAlaLeu 521
 Db 1603 TGGGATATGTACACT-----GATGAGGCAAGGAGATGCTTCAGATTCAATGATT 1653
 QY 522 GlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGly 534
 Db 1654 GGTAACCTATGAGAACTGACGACTTTAGAAATCGAGGA 1692

RESULT 11

AAQ10313 standard; DNA; 1692 BP.

ID AAQ10313
 AC AAQ10313;
 XX
 XX 24-OCT-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 05-APR-1991 (first entry)
 XX
 DE Sequence encoding protein with lipase activity.
 XX
 KW ATCC 34614; ds.
 XX
 OS Galactomyces geotrichum.
 XX
 FH Key
 FT CDS 1..1692
 FT /tag= a
 XX
 PN JP0229588-A.
 XX
 PD 11-DEC-1990.

XX 27-MAR-1989; 89JP-00074721.
 XX
 PR 27-MAR-1989; 89JP-00074721.
 XX
 PA (KURK) KURITA WATER IND LTD.
 PA (OSAK) OSAKA CITY.
 XX
 DR WPI; 1991-027567/04.
 XX P-PSDB; AAR10330.
 XX
 PT Gene for coding protein with lipase activity - is prepd. from messenger
 PT ribonucleic acid of geo-trichum candidum ATCC 34614.
 XX
 PS Claim 1; Fig 4; 12pp; Japanese.
 XX
 CC The gene product may be isolated from a transformed expression sytem, and
 CC may be enhanced with stability in heat, alkali, acid and organic solvent
 CC by position-specific modulation. (Updated on 25-MAR-2003 to correct PA
 CC field.) (Updated on 24-OCT-2003 to standardise OS field)
 XX
 SQ Sequence 1692 BP; 336 A; 470 C; 399 G; 487 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1-26e-94 Length: 1692
 Score: 1097.00 Matches: 238
 Percent Similarity: 56.16% Conservatives: 72
 Best Local Similarity: 43.12% Mismatches: 200
 Query Match: 38.42% Indels: 42
 DB: 2 Gaps: 8

US-09-943-857-4 (1-547) x AAQ10313 (1-1692)

QY 9 GlyArgLeuGlySerValProThrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeu 28
 DB 46 GGCACCTTGGCCAGCCGCCCGCTTCTTAATGGCAACGAGGTCTCTGTGTGC 105
 QY 29 AsnAlaIleAsnGluAlaPheLeuGlyIleProPheAlaGluProValGlyAsn 48
 DB 106 CTTGAGGGCAAGTGTACCTTCAAGGAAATCCCATTTGCTGACCTCTCTGTGTGAC 165
 QY 49 LeuArgPheLysAspProValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSer 68
 DB 166 TTGCGGTTCAAGCACCCCGAGCTTTCACTGGATCTTACAGGCTTAAAGGCCAACGAC 225
 QY 69 TyrGlyProSerCysMetGlnGlnAsnProGluGlyThrPheGlu----- 83
 DB 226 TTCAGCTCTGCTGTATGACGCTTGATCTCTGGCAATGCCATTTCTTTGCTTGACAAAGTC 285
 QY 84 -----GluAsnLeuGlyLysThrAlaLeuAspLeuValMet 95
 DB 286 GTGGGCTTGGGAAAGATTATTCCTGATACCTTAGAGGCCCTTTTATGACATGGCC--- 342
 QY 96 GlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsnVal 115
 DB 343 -----CAGGCTAGTGTTCATCAATGAGGACGTCTCTACCTTAACGTT 387
 QY 116 ValArgProGlyThrLysAlaGlyAlaAsnLeuProValMetLeuThrIlePheGly 135
 DB 388 TTCGCGCCCGCTGGCCCAAGCCTGATGCTAAGCTCCCGCTCATGGTTGGATTACGGT 447
 QY 136 GlyGlyPheGluIleGlySerProThrIlePheProAlaGlnMetValThrLysSer 155
 DB 448 GGTGCTCTGTGTGTGTCTTCTGCTTCTTACCTGGTAAACGCTTACGTTCAAGGAGAGT 507
 QY 156 ValLeuMetGlyLysHisIleIleHisValAlaValAsnTyrArgValAlaSerTrpGly 175
 DB 508 CTGGAATGGCCAGCCTGT 567
 QY 176 PheLeuAlaGlyAspAspIleLysAlaGluGlySerClyAsnAlaGlyLeuLysAspGln 195
 DB 568 TTCCTGGGTGTGATGCCATCACCGCTGAGGGTAACCAACCAACGCTGTGTCTGTGACGACCAG 627

QY 196 ArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLys 215
 DB 628 CGCAAGGGTCTCGAGTGGTGTAGCAACAACATTCGCACTTTGGTGGTGTATCCGACCAAG 687
 QY 216 ValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsn 235
 DB 688 GTCATGATTTTTCGGTGGTCCGCTGGTCCATGAGTGTGTCTACCACGCTTGTGTGCTTAT 747
 QY 236 AspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGly 255
 DB 748 GGTGTGTGACACACATCAACGGAAGCAGCTTTTCCACTCTGCCATCTTTCAGTCTGGC 807
 QY 256 AlaMetValProSerAspProValAspGlyThrTyrGlyAsnGluIleThrAspLeuPhe 275
 DB 808 GGTCTCTCTTCTTACTTCTGCTCTGCTGGTCCGAGAGTGCCTACACGAGATTT 867
 QY 276 ValSerSerAlaGlyCys--GlySerAlaSerAspLys-----LeuAlaCysLeuArg 292
 DB 868 GCTCATGATGCGGATGTGATGCCAGCCCGGTGACATGAATGAACCTCTGGCTTGTCTCCGC 927
 QY 293 SerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsn----- 305
 DB 928 AGCAAGTCCAGCGATGCTTGCACAGTGCACAGAACTCGTAGATCTCAAGGACCTGTT 987
 QY 306 -----ThrProGlyPheLeuAlaTyrSerSerLeuArgLeuSerTyrLeuProArgPro 323
 DB 988 GGTCTGCTCCCTCAATTCCTTGGATTGTT-----CCGAGACCC 1026
 QY 324 AspGlyLysAsnIleThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSer 343
 DB 1027 GACGGCAACATTATTCGCGATGCCCTTATGAACCTTACCGCAGCGGTAGATACGCCAAG 1086
 QY 344 ValProValIleIleGlyAspGlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeu 363
 DB 1087 GTTCTCTACATTACTGTAAACGAGGATGAGGTACTATTCTTCCGCCCTGGGTGCTATT 1146
 QY 364 AsnValThrThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPheIleHisAlaSer 383
 DB 1147 AATGTACACACTACTCCCATCTTAAAGAGTGTGTAAGTACATTTTGTAGCAGGGTCTT 1206
 QY 384 AspAlaGluIleAspThrLeuMetAlaLalaTyrProGlnAspIleThrGlnGlySerPro 403
 DB 1207 GACGCTTGCCTGTGATGCTGTTTGTCTTACCCCGGCTCTTGTGCGAGGGTGGCCA 1266
 QY 404 PheAspThrGlyValLeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeu 423
 DB 1267 TTCGCACTGGCATCTCAATGCTCTGACCCCTCAATTCAAGCGCATTTGCTGCAATTTTC 1326
 QY 424 GlyAspLeuAlaPheIleHisAlaArgArgTyrPheLeuAsnHisPheGlnGlyGlyThr 443
 DB 1327 ACTGATTTTCTGTTCCAGTCTCTCTGCTGTATGCTTAAAGCTTAAAGGACGTCAC 1386
 QY 444 LysTyrSerPheLeuSerLysGlnLeuSerGlyLeu---ProIleMetGlyThrPheHis 462
 DB 1387 CGCTGAGTACTTCTTCCCAATCACTACGCTGGACCTTGGTCTCTCTTCTGCTTCCAT 1446
 QY 463 AlaAsnAspIleValTrpGlnAspTyrLeuLeuGlySerGlySerValIleTyrAsnAsn 482
 DB 1447 GGTAGTACTTCTTCTTCCCAATCACTACGCTGGACCTTGGTCTCTCTTCTGCTTCC 1506
 QY 483 AlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAlaGlyLeuLeuValAsnTrp 502
 DB 1507 TACTTTATCTCGTTTCCCAACCAACGACGACCCCAACGTTGGCCACCACTGAACAGTGG 1566
 QY 503 ProLysTyrThrSerSerGlnSerGlyAsnAsnLeuMetMetIleAsnAlaLeuGly 522
 DB 1567 GATATGTACT-----GATTTCAGCAAGAGAGATGCTTCAGATTCATATGATGTT 1617
 QY 523 LeuTyrThrGlyLysAspAsnPheArgThrAlaGly 534
 DB 1618 AACTCTATGAGAACTCAGCACTTTAGAAATCGAGGA 1653


```

Db 1168 GTTTGTGCTCTACCCGGCTCTTGTGCGAGGTCGCCATTCGCACTGGTATTCTT 1227
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 410 AsnAlaLeuThrProGlnPheLysArgIleSerValLeuGlyAspLeuAlaPheIle 429
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1228 AATGCTCTGACCCCTCAGTTCAGCGCATTCCTGCCATTTTTCACCTGATTTGCTGTCAG 1287
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 430 HisAlaArgArgTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSer 449
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1288 TCTCCTCGTCTGTATGCTTTAAGCTACCAAGGACGTCAACCGCTGGACTTACTTGCC 1347
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 450 LysGlnLeuSerGlyLeu---ProIleMetGlyThrPheHisAlaAsnAspIleValTrp 468
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1348 ACCCAGCTCCATAACCTCGTTCCATTTTGGGTACTTTCCATGGTGTAGTGAICTCTTTTC 1407
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 469 GlnAspTyrLeu---LeuGlySerGlySerValIleTyrAsnAsnAlaPheIleAlaPhe 487
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1408 CAATACTACGTGGACCTTGGCCCATCTTCTGCT---TACCCTCGCTACTTTTATCTCGTTT 1464
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 488 AlaThrAspLeuAspProAsnThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSer 507
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1465 GCCAACCCACGACCCCAACGCTGGCACCACCTGAACAGTGGGATATGTACACT--- 1521
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 508 SerSerGlnSerGlyAsnAsnLeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLys 527
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1522 -----GATGCAGGCAAGGAGATGCTTCAGATTTCATATGGTTGGTAACTCTATGAGAACT 1575
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 528 AspAsnPheArgThrAlaGly 534
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1576 GACGACTTTAGAAATCGAGGGA 1596
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

```

RESULT 13

AAQ05605 standard; cDNA; 1674 BP.

XX ID

XX AC

XX DT

XX DT

XX DT

XX DE

XX KW

XX OS

XX FH

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

CC transform hosts for the prodn. of a protein with lipase activity. See
 CC also AAQ05605. (Updated on 25-MAR-2003 to correct PA field.) (Updated on
 CC 24-OCT-2003 to standardise OS field)

XX Sequence 1674 BP; 351 A; 473 C; 381 G; 469 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 8.08e-93 Length: 1674
 Score: 1078.00 Matches: 233
 Percent Similarity: 56.43% Conservative: 83
 Best local Similarity: 41.61% Mismatches: 202
 Query Match: 37.64% Indels: 42
 DB: 2 Gaps: 8

US-09-943-857-4 (1-547) x AAQ05605 (1-1674)

Qy 8 AlaGlyArgLeuGlySerValProThrAlaLysLeuAlaAsnGlyAspThrIleThrGly 27
 Db 25 GCTGGCGTCTGGCCGACGCGCCGCTCTCTCAATGGCAAGAGTCACTCTCTGCT 84
 Qy 28 LeuAsnAlaIleIleAsnGluAlaPheLeuGlyIleProPheAlaGluProProValGly 47
 Db 85 GTCCTTGAGGCAAGGTGTATACCTTCAAGGGAATCCATTTGCTGACCCCTCTTGAAT 144
 Qy 48 AsnLeuArgPheLysAspProValProTyrSerGlySerLeuAsnGlyGlnLysPheThr 67
 Db 145 GACTTGGATTCAAGCACCCCGACGCTTTCATGATCTCTACAGGCTCTTAAAGCCCAAT 204
 Qy 68 SerTyrGlyProSerCysMetGlnGlnAsnProGluGlyThrPheGluGluAsnLeuGly 87
 Db 205 GATTTAGCCCTGCTTGTATGACGCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 249
 Qy 88 LysThrAlaLeuAspLeuValMet---GlnSerLysValPhe----- 100
 Db 250 ---ACTTTGCTTGACAAAGCTCTGGATTGCGCAAAAGTCAATCCCGAAGAAATTTAGAGT 306
 Qy 101 -----GlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsn 114
 Db 307 CCCCTTTATGATATATGCGGCAAGGTACCGTGTGATGAATGAGGACTGCTTTTACCTCAAT 366
 Qy 115 ValValArgProGlyThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePhe 134
 Db 367 GTTTTCGCGCTGCTGGCACCAGCTGATGCTAAGCTCCCGCTGCTGCTGCTGCTGCTGCTGCT 426
 Qy 135 GlyGlyGlyPheGluIleGlySerProThrIlePheProProAlaGlnMetValThrLys 154
 Db 427 GGTGTGCTGCTTGTTCACGTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 486
 Qy 155 SerValLeuMetGlyLysHisIleIleHisValAlaValAsnTyrArgValAlaSerTrp 174
 Db 487 AGTATCAACATGGGCGCAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 546
 Qy 175 GlyPheLeuAlaGlyAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAsp 194
 Db 547 GGATTCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 606
 Qy 195 GlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSer 214
 Db 607 CAGCGCAAGGCTCTCGAGTGGGTAGCGCAACATTCGCAACTTTGGTGGTGGTGGTGGTGGTGG 666
 Qy 215 LysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrp 234
 Db 667 AAGGTCAATGATTTTCGCTGAGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 726
 Qy 235 AsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSer 254
 Db 727 TATGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 786
 Qy 255 GlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGluIleTyrAspLeu 274
 Db 787 GGTGGCCCTCTCTTCTTACCACGACTCTAGTCCGTTGGTCCGATATTTCTTACACAGAGA 846
 Qy 275 PheValSerSerAlaGlyCys-----GlySerAlaSerAspLysLeuAlaCysLeu 291

Protein gene having lipase activity - has defined sequence of 544 bases
 and gives lipase producing vector on integration to expression
 (secretion) vector.

XX Disclosure; Fig 4; 12pp; Japanese.

XX The cloned cDNA can be inserted into an expression vector and used to

```

Db      847  TTTGCTCAGTATCGCGATGTCACACTAGTCGCCAGTCCCAACGACACTCTGGAGTGTCTC 906
QY      292  ArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsn----- 305
Db      907  CGCAGCAAGTCCAGCTGCTGCTGACGATGCCAGAACTCTTACGATCTCAAGATCTG 966
QY      306  -----ThrProGlyPheLeuAlaTyrSerSerLeuArgLeuSerTyrLeuProArg 322
Db      967  TTTGGTCTACTCCCTCAATTCCTTGGATTGGT-----CCAGA 1005
QY      323  ProAspGlyLysAsnLeuThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAla 342
Db      1006  CCCAGCGGCAACATTATTCGCGATGCCGCTATGAGCTCTTCGCGACGCTAGATACGCC 1065
QY      343  SerValProValIleLeuGlyAspGlnAsnAspGluGlyThrIlePheGlyLeuSerSer 362
Db      1066  AAGGTTCCCTACATTAGCGGTAAACCGAGAGATGAAGTACTGCTTTTGTCTCTGTGTG 1125
QY      363  LeuAsnValThrThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPheIleHisAla 382
Db      1126  CTCACGCTACCAAGCTCCCATGTTAAGAGTGGTTGAGTACATTTTCTACGATGCT 1185
QY      383  SerAspAlaGluIleAspThrLeuMetAlaTyrProGlnAspIleThrGlnGlySer 402
Db      1186  TCCGAGGCTTCCATTGACCGTGTGTTGTCGCTGTACCCGCGACACCTCTCTGTGTGCTG 1245
QY      403  ProPheAspThrGlyValLeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaVal 422
Db      1246  CCTTCGCACTGGCATCTTAATGCGCTACCCCGCTCAAGCGTGTGCGGCCATC 1305
QY      423  LeuGlyAspLeuAlaPheIleHisAlaArgTyrPheLeuAsnHisPheGlnGlyGly 442
Db      1306  TTGTCGATATGCTTTTCCAGTCTCCCGCGCGTGATGCTTAGCGCCACCAAGGAGTT 1365
QY      443  ThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeu---ProIleMetGlyThrPhe 461
Db      1366  AACCGCTGACATTACCTTTTCGACCCATCTGCACAACCTGTGTCATTTTGGGTACTTTC 1425
QY      462  HisAlaAsnAspIleValTrpGlnAspTyrLeuLeuGlySerGlySerValIleTyrAsn 481
Db      1426  CATGGCAACGAGCTTATCTTCCAAATCAATGAAACATTCGCGCCCGCTAACTCCTACCTT 1485
QY      482  AsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAlaGlyLeuLeuValAsn 501
Db      1486  CGTTATTTTATTTCCTTTGCGCAACCAACCATGACCTTAATGTTGGTACTAATCTGCTCCAG 1545
QY      502  TrpProLysTyrThrSerSerSerGlnSerGlyAsnAsnLeuMetMetIleAsnAlaLeu 521
Db      1546  TGGGATCAATACACT-----GATGAAGGCAAGGAGATGCTTGAGATTACATGACC 1596
QY      522  GlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGlyTyrAspAlaLeuMetThrAsn 541
Db      1597  GATAATGTCATGAGACTGATGACTACAGAAATTGAGGGAATCTCAAACTTTGAGACTGAC 1656

RESULT 14
ID      AAQ98578
XX      AAQ98578 standard; DNA; 2045 BP.
AC      AAQ98578;
XX
DT      05-MAR-1996 (first entry)
DE
XX      Aspergillus foetidus glucoamylase gene promoter.
XX      recombinant; glucose oxidase; GOD; extracellular production;
KW      filamentous fungus; glucoamylase; promoter; ss.
XX
OS      Aspergillus foetidus; SE4.
PN      EP665291-A1.
XX

```

```

PD      02-AUG-1995.
XX
PF      19-JAN-1995; 95EP-00200129.
XX
PR      28-JAN-1994; 94BE-00000102.
PR      17-JUN-1994; 94BE-00000586.
PR      09-JAN-1995; 95BE-00000014.
XX
PA      (SOLV ) SOLVAY & CIE.
XX
PI      Carrez D, Roos J;
DR      WPI; 1995-264864/35.
XX
PT      Expression system for extracellular prodn. of glucose oxidase - related
PT      vectors and transformed cells, esp. new strain of Aspergillus foetidus,
XX      provides high yields of enzyme uncontaminated by catalase.
PS      Claim 18; Page 17-19; 32pp; French.
XX
CC      The promoter and terminator sequences from the glucoamylase gene of
CC      Aspergillus foetidus SE4 strain (AAQ98578 and AAQ98570, respectively) are
CC      used for extracellular expression of glucose oxidase in A.foetidus hosts.
CC      The specifically claimed transformant strain SE4tr contains a vector in
CC      which the GOD gene is under transcriptional control of the new promoter
CC      and terminator sequences; extracellular production of GOD by the
CC      transformant was 5000-10000 times greater than for the untransformed
CC      parental SE4 strain. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ      Sequence 2045 BP; 454 A; 538 C; 565 G; 488 T; 0 U; 0 Other;

```

```

Alignment Scores:
Pred. No.: 2,16e-79 Length: 2045
Score: 938.50 Matches: 207
Percent Similarity: 60.81% Conservative: 80
Best Local Similarity: 43.86% Mismatches: 162
Query Match: 32.77% Indels: 24
DB: 2 Gaps: 8

```

US-09-943-857-4 (1-547) x AAQ98578 (1-2045)

```

QY      89  ThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuProGlnSerGlu 108
Db      13  TCGGTCAATTGGCTCTCTCGCTGATCTCCCTTTGGTACAGTCGCGCTACCAATGCTCTCGAG 72
QY      109  AspCysLeuThrIleAsnValValArgProProGlyThrLysAlaGlyAlaAsnLeuPro 128
Db      73  GATTGCTGAACATTGACATTGCGCGTCCGCGCGGACCAACCCGCGACTCGAAGCTGCCT 132
QY      129  ValMetLeuTriPhePheGlyGlyGlyPheGluIleGlySerProThrIlePheProPro 148
Db      133  GTGCTGGTCTGGATCTTTGGCGGAGCTTTGAACCTTGTTTCAAGCGGATGATGATGTT 192
QY      149  AlaGlnMetValThrLysSerValLeuMetGlyLysHisIleIleHisValAlaValAsn 168
Db      193  ACAACGATGGTATCATCGTCGATAGACAAGAACATCGCTATCGTGTGTTAGCAATGAAT 252
QY      169  TyrArgValAlaSerTrpGlyPheLeuAlaGlyAspAspIleLysAlaGluGlySerGly 188
Db      253  TATCGCGTGGGAGGTTTCGGGTTCTTCCCGGAAGGAGATCCTCGAGGACGGGTCCGCG 312
QY      189  AsnAlaGlyLeuLysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGly 208
Db      313  AACCTAGGGCTCTCTGGACCAACGCCCTTGCCCTGCGAGTGGGTTGCCACAACTCGAGGCC 372
QY      209  PheGlyGlyAspProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerVal 228
Db      373  TTTGGTGGAGACCCGACCAAGGTGACGATTTGGGGAGAAATCAGCAGGAGCCATTCGTTTG 432
QY      229  -LeuCysHisLeuIleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheAr 248
Db      433  ACTAGATGACTTGTGAC-----GACGGAAACATCACTTACAAGGATAAGCCCTTGTTCGG 486

```

QY 248 gAlaGlyIleMetGlnSerGlyValMetValProSerAspProValAspGlyThrTyrGl 268
 Db 487 GGGGGCCATCATGACTCCGCTAGTGTGTTCCCGCAGACCCCGTCGATGGGTCAAGG 546
 QY 268 yAsnGluIleTyrAspLeuPheValSerSerAlaGlyCysGlySerAlaSerAspLysLe 288
 Db 547 ACAGCAAGTATATGATGGGTAGTGAATCTGCAGGCTGTCTCTTCTTAACGACACCT 606
 QY 288 uAlaCysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGl 308
 Db 607 AGCTGTCTGGTGAACCTAGACTACACCGACTTCCTCAATGGCGAAATCCCGTCCAGG 666
 QY 308 yPheLeuAlaTyrSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIl 328
 Db 667 CATTTAAGCTACCATCTGTGGCTTATCATATGTGCTCGACCGGACGGCGGCTT 726
 QY 328 eThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleIl 348
 Db 727 GTCGGCATCACCGGACGTTTGGGCAAGACGAGGAATATGCTCGGCTCCGCTTCATCGT 786
 QY 348 eGlyAspGlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAs 368
 Db 787 GGGCACCACAGAGGATAGGGACCTTATTCGCTTGTTCAGTCCAACTTACGACAT 846
 QY 368 nAlaGlnAlaArgAlaTyrPheLysGlnSerPheIleHis---AlaSerAspAlaGluIl 387
 Db 847 CGACGAGTGTGCTGACTACCTGGCTCATCTTCTTATGACGCTAGCGGAGACGCT 906
 QY 387 eAspThrLeuMetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThrG 407
 Db 907 TGAAGAACTAGTGGCCCTGTATCCACAGACACCACGCTACGCGGTCTCGCTTCAGACGG 966
 QY 407 lyValLeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuA 427
 Db 967 CGGCC---AACACTGGTATCCGCAATTGAAGCAATTGGCCCAATCTCGCGCACTGG 1023
 QY 427 laPheIleHisAlaArgArgTyrPheLeuAsnHisPheGlnGlyGlyThr----- 443
 Db 1024 TCTTCAC-CATTAACCGGGGCACTTCTCTGATGCAGAGGAATCTCCCTGATCTTC 1082
 QY 444 --LysTyrSerPheLeuSerGlnLeuSerGlyLeuProIleMetGlyThrPheHisA 463
 Db 1083 CGAACTGGTGTGCTGACCTGGCGACCTATGACTATGGCACCCTTCTCTGGGACCTTCCACG 1142
 QY 463 laAsnAspIleValTrpGlnAspTyr-----LeuLeuGlySerGlySerV 478
 Db 1143 GAAGTGACCTGCTCGAGGTGTTCTATGGATCGACGCCAACTATCGAGTAGTCTTAGCC 1202
 QY 478 alIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThr---AlaG 497
 Db 1203 ACACGTAC-----TATCTGAGCTTTGTATAGCTGATCGGATCCGAACCTCAACCGGG 1253
 QY 497 lyLeuLeuValAsnTrpProLysTyrThrSerSerSerGlnSerGlyAsnAsnLeuMetM 517
 Db 1254 GGGAGTACATTAGTGGCGCGCAGTGGAGGAATCGCGGCG-----TTGATGA 1301
 QY 517 erIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGlyTyrAspA 537
 Db 1302 ATTTGGGCGGACGAGCGGCTCTCTTACGATGATTTCCGCAACCGGACATATGAT 1361
 QY 537 laLeuMetThrAsnProSerSerPhePheVal 547
 Db 1362 TCATCTCGCAGAAATACCGGCGGCTTCCACATC 1393
 RESULT 15
 ID AAT11299
 XX AAT11299 standard; DNA; 2045 BP.
 AC AAT11299;
 XX
 XX
 DT 16-OCT-2003 (revised)
 DT 25-JUL-1996 (first entry)
 XX

DE Aspergillus foetidus SE4 glucoamylase promoter sequence.
 XX
 KW Microbial aspartic protease; filamentous fungus; aspergillopepsin;
 KW extracellular production; secretion; glucoamylase gene promoter; ss.
 XX
 OS Aspergillus foetidus; SE4.
 XX
 PN EP687734-A1.
 XX
 PD 20-DEC-1995.
 XX
 PF 09-JUN-1995; 95EP-00201521.
 XX
 PR 17-JUN-1994; 94BE-00000585.
 PR 09-JAN-1995; 95BE-00000015.
 XX
 PA (SOLV) SOLVAY SA.
 XX
 PI Carrez D, Dhaese P;
 XX
 DR WPI; 1996-031780/04.
 XX
 PT System for extracellular prodn. of microbial aspartic protease - contg.
 PT promoter, signal sequence, pro-peptide and mature enzyme sequences, and
 PT terminator, esp. for aspergillopepsin prodn.
 XX
 PS Claim 19; Page 18-20; 35pp; French.
 XX
 CC Spores of A. foetidus ATCC 14916 were exposed to UV, then grown on a
 CC nutrient medium and one mutant having increased production of
 CC glucoamylase was selected and designated SE4. The promoter and terminator
 CC regions of the glucoamylase gene from mutant strain SE4 were isolated and
 CC sequenced. These transcription control regions are preferred for
 CC incorporation into novel expression vectors designed to allow
 CC extracellular production of microbial aspartic protease. In particular,
 CC the enzyme aspergillopepsin is produced by transformed filamentous fungi
 CC and secreted into the culture medium due to the presence of a fungal
 CC secretion peptide. The present sequence is that of the A. foetidus SE4
 CC glucoamylase promoter. (Updated on 16-OCT-2003 to standardise OS field)
 XX
 SQ Sequence 2045 BP; 454 A; 538 C; 565 G; 488 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 2,16e-79 Length: 2045
 Score: 938.50 Matches: 207
 Percent Similarity: 60.81% Conservative: 80
 Best Local Similarity: 43.86% Mismatches: 162
 Query Match: 32.77% Indels: 24
 DB: 2 Gaps: 6
 US-09-943-857-4 (1-547) x AAT11299 (1-2045)
 QY 89 ThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuProGlnSerGlu 108
 Db 13 TCGGTGTCATGGCTCTCTCGCTGATCTCCCTTTGTGACAGTCGCTACCAATGCTCTCGAG 72
 QY 109 AspCysLeuThrIleAsnValValArgProProGlyThrLysAlaGlyAlaAsnLeuPro 128
 Db 73 GATTGCTGAACTTGCATTCGGCGTCCGGCGGAGCCACCGCGAGCTCGAAGTGCCT 132
 QY 129 ValMetLeuTrpIlePheGlyGlyPheGluIleGlySerProThrIlePheProPro 148
 Db 133 GTGCTGTGCTGATCTTTGGCGGAGGCTTTGAACCTTGGTTCAAAGCGGATGATGATGCT 192
 QY 149 AlaGlnMetValThrLysSerValLeuMetGlyLysHisIleIleHisValAlaValAsn 168
 Db 193 ACAACGATGATCATCTGCTGATACACAGACATGCTGCTGTTGTGACCATGAT 252
 QY 169 TyrArgValAlaSerTrpGlyPheLeuAlaGlyAspAlleLysAlaGluGlySerGly 188
 Db 253 TATCGCGTGGGAGGTTTCGGGTTCCTGCCGGAAGAGAGATCTCGGAGCGGTCGCGG 312
 QY 189 AsnAlaGlyLeuLysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGly 208

B/GNK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 5, 2004, 19:05:41 ; Search time 91 Seconds
(without alignments)
3335.804 Million cell updates/sec

Title: US-09-943-857-4
Perfect score: 2864
Sequence: 1 SMNSRGPAGRLGVSPTAKLA.....DNFRTAGYDALMTNPSFFV 547

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DB=xlh

-MODEL=frame p2n.model -DEV=xlh
-O=/cgn2_1/USPTO spool/US09943857/runat_29072004_090901_29773/app_query.fasta.1.711
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rnl -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09943857 @CGN 1.1 54 @runat_29072004_090901_29773 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/2/ina/5A COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTCUTS COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2415	84.3	1650	2	US-08-776-210-6
2	1107	38.7	1928	3	US-08-362-525-11
3	1089	38.0	1635	2	US-08-776-210-5
4	938.5	32.8	2045	1	US-08-379-928A-7
5	632	22.1	8533	1	US-07-846-181-6
6	632	22.1	8533	1	US-07-845-989-6
7	452.5	15.8	1738	2	US-08-932-376A-1
8	452	15.8	1845	1	US-07-732-962A-1
9	452	15.8	1845	5	PCT-US92-06106-1
10	452	15.8	2256	2	US-08-318-826A-5
11	452	15.8	2256	2	US-08-370-156-1
12	452	15.8	2256	3	US-08-814-095-1

13	452	15.8	2400	6	5215909-13	Patent No. 5215909
14	452	15.8	3016	2	US-08-318-826A-7	Sequence 7, Appli
15	452	15.8	3016	2	US-08-370-156-5	Sequence 5, Appli
16	452	15.8	3016	3	US-08-814-095-5	Sequence 5, Appli
17	452	15.8	3096	2	US-08-318-826A-6	Sequence 6, Appli
18	452	15.8	3096	2	US-08-370-156-3	Sequence 3, Appli
19	452	15.8	3096	3	US-08-814-095-3	Sequence 3, Appli
20	451	15.7	2381	2	US-08-318-826A-9	Sequence 9, Appli
21	451	15.7	2416	2	US-08-318-826A-8	Sequence 8, Appli
22	451	15.7	2416	3	US-09-334-489-1	Sequence 1, Appli
23	451	15.7	2416	3	US-09-334-489-2	Sequence 2, Appli
24	444.5	15.5	2445	6	5215909-9	Patent No. 5215909
25	437.5	15.3	1470	1	US-08-589-893-17	Sequence 17, Appl
26	437.5	15.3	1470	2	US-09-020-991-17	Sequence 17, Appl
27	437.5	15.3	1470	2	US-09-062-890-17	Sequence 17, Appl
28	436.5	15.2	1470	1	US-08-589-893-11	Sequence 11, Appl
29	436.5	15.2	1470	2	US-09-062-890-11	Sequence 11, Appl
30	436.5	15.2	1470	2	US-09-062-890-11	Sequence 11, Appl
31	432.5	15.1	1470	1	US-08-589-893-3	Sequence 3, Appli
32	432.5	15.1	1470	1	US-08-589-893-5	Sequence 5, Appli
33	432.5	15.1	1470	1	US-08-589-893-19	Sequence 19, Appl
34	432.5	15.1	1470	2	US-09-020-991-3	Sequence 3, Appli
35	432.5	15.1	1470	2	US-09-020-991-3	Sequence 3, Appli
36	432.5	15.1	1470	2	US-09-020-991-19	Sequence 19, Appl
37	432.5	15.1	1470	2	US-09-062-890-3	Sequence 3, Appli
38	432.5	15.1	1470	2	US-09-062-890-5	Sequence 5, Appli
39	432.5	15.1	1470	2	US-09-062-890-19	Sequence 19, Appl
40	432.5	15.1	1470	2	US-09-062-890-27	Sequence 27, Appl
41	431.5	15.1	1470	1	US-08-124-674-1	Sequence 1, Appli
42	431.5	15.1	1470	1	US-08-589-893-1	Sequence 1, Appli
43	431.5	15.1	1470	2	US-09-020-991-1	Sequence 1, Appli
44	431.5	15.1	1470	2	US-09-062-890-1	Sequence 1, Appli
45	430.5	15.0	1470	1	US-08-589-893-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-08-776-210-6
; Sequence 6, Application US/08776210
; Patent No. 5942659
; GENERAL INFORMATION:
; APPLICANT: ALIBERT, Gilbert
; APPLICANT: MOULOUNGUI, Zephirin
; APPLICANT: BOUDET, Alain
; TITLE OF INVENTION: PROCESS FOR PRODUCING FATTY ACIDS OR
; DERIVATIVES THEREOF FROM OLEAGINOUS PLANTS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSER: YOUNG & THOMPSON
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/776,210
; FILING DATE: 24-JAN-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 94 09272
; FILING DATE: 25-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR95/00957
; FILING DATE: 18-JUL-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J
; REGISTRATION NUMBER: 32,925

REFERENCE/DOCKET NUMBER: IN 387 - BE 6996

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-521-2297

TELEFAX: 703-685-0573

TELEX: 248425 EMBON

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 1650 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cdna

US-08-776-210-6

Alignment Scores:

Pred. No.: 3,86e-262 Length: 1650
Score: 2415.00 Matches: 454
Percent Similarity: 90.50% Conservative: 32
Best Local Similarity: 84.54% Mismatches: 51
Query Match: 84.32% Indels: 0
DB: 2 Gaps: 0

US-09-943-857-4 (1-547) x US-08-776-210-6 (1-1650)

```

Qy 11 LeuGlySerValProThrAlaLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAla 30
Db 37 GTGGCTGCTGCCCCACCGCCACGCTGCCAACGGCGACACCATCATCAGCGTCTCAAGCC 96
Qy 31 IleIleAsnGluAlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArg 50
Db 97 ATCATCAAGAGGCGTCTCTCGGCATTCCCTTTGCCGAGCGCGGTGGGCAACCTCCGC 156
Qy 51 PheLysAspProValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGly 70
Db 157 TTCAAGGACCCCGTCTACTCCGGCTCGCTCGATGCGCCAGAGTTCACGCTGTACGCG 216
Qy 71 ProSerCysMetGlnGlnAsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAla 90
Db 217 CCGTGTGCATGACAGAACCCCGAGGCGACCTACGNGAGAGACCTCCCAAGCGCAGG 276
Qy 91 LeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCys 110
Db 277 CTCGACTTGTGTGATGCACTCAAGGTGTTGAGCGCGTGTCTGCGCTGAGCGGAGACTGT 336
Qy 111 LeuThrIleAsnValValArgProGlyThrLysAlaGlyAlaAsnLeuProValMet 130
Db 337 CTCACCATCAAGCTGTGCGCGCGCGGCGCAAGCGGGTGGCCAACTCCCGGTGATG 396
Qy 131 LeuTrpIlePheGlyGlyPheGluIleGlySerProThrIlePheProAlaGln 150
Db 397 CTCTGGATCTTTGGCGCGGGTTGAGGTGGTGCCACGACCACTTCCCTCCCGCCGAG 456
Qy 151 MetValThrLysSerValLeuMetGlyLysHisIleIleHisValAlaValAsnTyrArg 170
Db 457 ATGATCACCAAGAGCATTTGCCATGGGCAAGCCCATCATCCACGTGAGCGTCAATACCG 516
Qy 171 ValAlaSerTrpGlyPheLeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsnAla 190
Db 517 GTGTGCTGTGGGGTCTTGTGGCTGCGCAGCATCAAGGCGGAGGCGAGTGCACCGCC 576
Qy 191 GlyLeuLysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGly 210
Db 577 GGTTTGAAGGACACCGCTTGGCATGCGTGGTGGCGGACAAATTGCGCGGTTTGGC 636
Qy 211 GlyAspProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCys 230
Db 637 GCGGACCCGACCAAGGTGACCATCTTTGGCGAGCTGGCGGCGAGCATGTGCGTCAATGTC 696
Qy 231 HisLeuIleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGly 250
Db 697 CACATCTCTGGAACGACGCGGACAAACGTAAGGGCAAGCCGCTCTTCCCGCGGCG 756
Qy 251 IleMetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGlu 270

```

```

Db 757 ATCATGACGCTGGGGCCATGTGTCCGCTGGACGCCCTGGACGGCATCTACGGCAACGAG 816
Qy 271 IleTyrAspLeuPheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCys 290
Db 817 ATCTTTGACCTCTTGGCGTGAACGCGGTGGGAGCGCCGACGACAGCTTGCCTGC 876
Qy 291 LeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeu 310
Db 877 TTGCGCGGTGTGCTGAGGACACGTTGGAGGACGCCACCAACAACACCCCTGGTCTTG 936
Qy 311 AlaTyrSerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThrAsp 330
Db 937 GCCTACTCTCTGCTTGGCGTGTGTACTCCCGCGCGCCGACGCGTGAACATCACCGAC 996
Qy 331 AspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleIleGlyAsp 350
Db 997 GACATGTACCCCTTGTGTGCGGAGGCAAGTATGCCAACATCCCTGTGTGATCATCGCG 1056
Qy 351 GlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnAlaGln 370
Db 1057 CAGAACGACGAGGCGACCTTCTTTGGCACCTGTGTGAACGTGACACGCGATGCCAG 1116
Qy 371 AlaArgAlaTyrPheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeu 390
Db 1117 GCCCGCGAGTACTTCAAGCAGCTGTTTGTCCACGCCGACGCGGAGATCGACACGCTTG 1176
Qy 391 MetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsn 410
Db 1177 ATGACGCGGTACCCCGCGGACATCACCCAGGCGCTGCGGTTCGACACGGGTATTCTCA 1236
Qy 411 AlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHis 430
Db 1237 GCCCTACCCCGCGAGTTCAAGAGAACTCTGCGCGTGTCTCGCGACCTTGGCTTTACGCT 1296
Qy 431 AlaArgArgTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSerLys 450
Db 1297 GCTCGTGCCTACTTCTCAACCACTACACCGCGCGGACCAAGTACTCATCTCTCTGAAG 1356
Qy 451 GlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAsp 470
Db 1357 CAGCTCTGGCGTTCGCGGTGCTCGGAACGTTCCACTCCAACGACATTTGTCTCCAGAC 1416
Qy 471 TyrLeuLeuGlySerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAsp 490
Db 1417 TACTTGTGGGACGCGCTCGCTCATCTACAAACAACGCGTTCAITTCGCTTTGCCACG 1476
Qy 491 LeuAspProAsnThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerGln 510
Db 1477 TTGACCCCAACACCGCGGGTGTGTGTGAAGTGCGCGGAGTACACGACGCGCTGCAG 1536
Qy 511 SerGlyAsnAsnLeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPhe 530
Db 1537 CTGGGCAACAACCTTGATGATCATCAACGCTTGGGCTGTGTACACCGCGCAAGCAACTTC 1596
Qy 531 ArgThrAlaGlyTyrAspAlaLeuMetThrAsnProSerSerPhePheVal 547
Db 1597 GCGACCGCGGCTACGACGCGTGTGTCTTCCAACCGCGCTGTCTTTGTG 1647

```

RESULT 2

US-08-362-525-11

; Sequence 11, Application US/08362525

; Patent No. 6027910

; GENERAL INFORMATION:

; APPLICANT: KLIS, FRANCISCUS M.

; APPLICANT: SCHREUDER, MAARTEN P.

; APPLICANT: TOSCHKA, HOLSER Y.

; APPLICANT: VERRIPS, CORNELIS T.

; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE

; TITLE OF INVENTION: CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION

; TITLE OF INVENTION: PROTEIN

; NUMBER OF SEQUENCES: 26

; CORRESPONDENCE ADDRESS:


```

384 AspAlaGluIleAspThrLeuMetAlaAlaTyrProGlnAspIleThrGlnGlySerPro 403
1246 GACGCTTCGCTGTATCGTGTTCCTGCTTACCCCGCTCTTGGTCGAGGGTTCACCA 1305
404 PheAspThrGlyValLeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeu 423
1306 TTCGCACTGGTATCTTAATGCTCTTACCCCTCAGTTCAGCGCAATGCTGCCATTTTC 1365
424 GlyAspLeuAlaPheIleHisAlaArgArgTyrPheLeuAsnHisPheGlnGlyThr 443
1366 ACTGATTTGCTGTTCCAGTCTCCTCGTGTGTTATGCTTAACGCTACCAAGCAGCTCAAC 1425
444 LysTyrSerPheLeuSerLysGlnLeuSerGlyLeu---ProIleMetGlyThrPheHis 462
1426 CGCTGGACCTTACCTTGCACCCAGCTCCATCACTTCGTTCCATTTTGGGTACTTCCAT 1485
463 AlaAsnAspIleValTrpGlnAspTyrLeu---LeuGlySerGlySerValIleTyrAsn 481
1486 GGCAGTGATCTCTTTTCAATACTACTACGTGACCTTGGCCCATCTTCTGCT---TACCGC 1542
482 AsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAlaGlyLeuLeuValAsn 501
1543 CGCTACTTTATCTCGTTTGGCAACACCAACGACCCCAACGCTTGGTACCACTCCACACAG 1602
502 TrpProLysTyrThrSerSerSerGlnSerGlyAsnAsnLeuMetMetIleAsnAlaLeu 521
1603 TGGGATATGACACT-----CATGAGCAAGAGATGCTTCAGATTTCATATGATT 1653
522 GlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGly 534
1654 GGTAACTCTATGAGAACTGACGACTTTAGAAATCGAGGGA 1692

```

RESULT 3

```

US-08-776-210-5
; Sequence 5, Application US/08776210
; Patent No. 5942659
; GENERAL INFORMATION:
; APPLICANT: ALIBERT, Gilbert
; APPLICANT: MOULOUNGUI, Zephirin
; APPLICANT: BOUDET, Alain
; TITLE OF INVENTION: PROCESS FOR PRODUCING FATTY ACIDS OR
; TITLE OF INVENTION: DERIVATIVES THEREOF FROM OLEAGINOUS PLANTS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSER: YOUNG & THOMPSON
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/776,210
; FILING DATE: 24-JAN-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 94 09272
; FILING DATE: 25-JUL-1994
; PRIOR APPLICATION DATA: WO PCT/FR95/00957
; APPLICATION NUMBER: 18-JUL-1995
; FILING DATE: 18-JUL-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J
; REGISTRATION NUMBER: 32,925
; REFERENCE/DOCKET NUMBER: IN 387 - BE 6996
; TELEPHONE: 703-521-2297
; TELEFAX: 703-685-0573
; TELEX: 248425 EMBON

```

```

; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1635 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-776-210-5

Alignment Scores:
Pred. No.: 1,25e-112 Length: 1635
Score: 1089.00 Matches: 238
Percent Similarity: 56.86% Conservative: 73
Best Local Similarity: 43.51% Mismatches: 192
Query Match: 38.02% Indels: 44
DB: 2 Gaps: 10

US-09-943-857-4 (1-547) x US-08-776-210-5 (1-1635)
Qy 15 ProThrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAlaIleAsnGlu 34
Db 7 CCCACGGCGCTTCTTAATGGCAACGAGGTCACTCTGCTGCTTGGGGGCAAGTTGAT 66
Qy 35 AlaPheLeuGlyIleProPheAlaGluProValGlyAsnLeuAsnLeuAspPheLysAspPro 54
Db 67 ACCTTTAAAGGAATTCCTGCTGACCTCTCTGTTGGTACTGCGGTTTCAGCACCCC 126
Qy 55 ValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyProSerCysMet 74
Db 127 CAGCTTTCACTGGATCCTACCGAGGTCTTAAGGCCAACGACTTCAGCTCTGCTGTATG 186
Qy 75 GlnGlnAsnProGluGlyThrPheGlu----- 83
Db 187 CAGCTTGATCTCTGGCAATGCCATTTCTTGGCTTGCAAAAGTCGTGGGCTTGGGAAAGATT 246
Qy 84 -----GluAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGln 101
Db 247 CTTCTGTATACCTTAGAGCGCCCTCTTTATGACATGGCC-----CAG 288
Qy 102 AlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgProGlyThr 121
Db 289 GGTAGTGTCTCCATGAATGAGGACGTCTCTACCTTAACGTTTCCGCGCTCTGTCGACC 348
Qy 122 LysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyGlyGlyPheGluIleGly 141
Db 349 AAGCCTGTATGCTAAGCTCCCGCTCATGTTTACGTTTACGGTGGTGGCTTTGTGTTGT 408
Qy 142 SerProThrIlePheProProAlaGlnMetValThrLysSerValLeuMetGlyLysHis 161
Db 409 TCTTCTGTCTTACCCTGGTAACGGCTACGTCAAGGAGAGTGTGGAATGGGCCAGCCT 468
Qy 162 IleIleHisValAlaValAsnTyrArgValAlaSerTyrGlyPheLeuAlaGlyAspAsp 181
Db 469 GTTGTGTTTGTTCATCACTACCTACCGCGCCCTATGATGTTCTGGTGTGTGATGCC 528
Qy 182 IleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMetGlnTrp 201
Db 529 ATCACCGCTGAGGGTAACCAACCAACGCTGTCTGCACGACGAGGAGGCTCTCGAGTGG 588
Qy 202 ValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysValThrIlePheGlyGlu 221
Db 589 GTTAGCGCAACATGCCAATTTGGTGTGTATCCCGACACAGGTTCATGATTTTCGTTGAG 648
Qy 222 SerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAspGlyAspAsnThrTyr 241
Db 649 TCCGCTGTTGCCATGATGTTGCTCACCAGCTTGTTCCTTACGCTGCTGTCGACACACCTAC 708
Qy 242 LysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValProSerAsp 261
Db 709 AACGAAAGAAAGCTTTTCCACTCTGCATCTTCAGTCTGCGCGCTCTCTCTCTTACTTT 768
Qy 262 ProValAspGlyThrTyrGlyAsnGluIleTyrAspLeuPheValSerSerAlaGlyCys 281

```

```

Db 769 GACTCTACTTCTGTGTGTCGAGAGTGCCTACAGCAGATTGTGCTCAGTATGCGGATGT 828
Qy 282 ---GlySerAlaSerAspLys-----LeuAlaCysLeuArgSerAlaSerSerAspThr 298
Db 829 GATGCCAGCGCCAGTGAACAATCTGCTGTCTCCGACGAAATGTCAGGATGTC 888
Qy 299 LeuLeuAspAlaThrAsnAsn-----ThrProGlyPhe 309
Db 889 TTGCACAGTCCCGAAGACTCTGACGATCTCAAGGACCTGTTTGGCCTGCTCCCTCAATTC 948
Qy 310 LeuAlaTyrSerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThr 329
Db 949 CTTGGATTGGT-----CCGAGACCCGACGCAACATTATTATCC 987
Qy 330 AspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleIleGly 349
Db 988 GATGCCGCTTATGAGCTCTACCGCAGCGGTAGATACGCCAAGTTCCTACATTACTGGT 1047
Qy 350 AspGlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnAla 369
Db 1048 AACCCAGGAGATGAGGATTAATCTTGTCCCGCTTGTCTTAATGCTACACGACTCC 1107
Qy 370 GlnAlaArgAlaTyrPheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThr 389
Db 1108 CATGTTAAGAGTGGTGAAGTACATTGTAGCGAGGCTTCTGACGCTTCGCTTGATCGT 1167
Qy 390 LeuMetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeu 409
Db 1168 GTTTGTGCGCTTACCCCGGCTCTGTGTGCGAGGTCGCCATTCCGACATGGTATCTT 1227
Qy 410 AsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIle 429
Db 1228 AATGCTCTGACCCCTCAGTTCAGCGCATGCTGCCATTTCATGATTTGCTGTTCCAG 1287
Qy 430 HisAlaArgArgTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSer 449
Db 1288 TCTCTCGTGTGTTATGCTTAACGTTACCAAGACGTCACCGCTGAGCTTACCTTGCC 1347
Qy 450 LysGlnLeuSerGlyLeu---ProIleMetGlyThrPheHisAlaAsnAspIleValTrp 468
Db 1348 ACCAGCTCCATAACCTCGTTTCATTTTTGGGTACTTTCATGTGTAGTATCTCTTTTC 1407
Qy 469 GlnAspTyrLeu---LeuGlySerGlySerValIleTyrAsnAsnAlaPheIleAlaPhe 487
Db 1408 CAATACTAGTGCACCTTGCCCATCTCTCTGCT---TACCGCGCTACTTTATCTCGTTT 1464
Qy 488 AlaThrAspLeuAspProAsnThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSer 507
Db 1465 GCCAACCCACGACCCCAACGTTGGCACCACCACTGAACAGTGGGATATGTACACT--- 1521
Qy 508 SerSerGlnSerGlyAsnAsnLeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLys 527
Db 1522 -----GATGCAGGCAAGGAGATGCTTCAGATTTCATATGTTGTGTAACTATGAGAACT 1575
Qy 528 AspAsnPheArgThrAlaGly 534
Db 1576 GACGACTTTAGATCGAGGGA 1596

```

RESULT 4

```

US-08-379-926A-7
; Sequence 7, Application US/08379926A
; Patent No. 5783414
; GENERAL INFORMATION:
; APPLICANT: CARREZ, DIRK
; APPLICANT: ROOS, JOEL
; TITLE OF INVENTION: EXPRESSION SYSTEM, INTEGRATION
; TITLE OF INVENTION: VECTOR
; TITLE OF INVENTION: AND CELL TRANSFORMED BY THIS INTEGRATION VECTOR
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER &
; ADDRESSEE: NEUSTADT
; STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400

```

```

; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/379,926A
; FILING DATE: 27-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: BE 09400102
; FILING DATE: 28-JAN-1994
; APPLICATION NUMBER: BE 09400586
; FILING DATE: 17-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: BE 09500014
; FILING DATE: 09-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 3987-13-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2045 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-379-926A-7

```

```

Alignment Scores:
Pred. No.: 1,69e-95 Length: 2045
Score: 938.50 Matches: 207
Percent Similarity: 60.81% Conservative: 80
Best Local Similarity: 43.86% Mismatches: 162
Query Match: 32.77% Indels: 24
DB: 1 Gaps: 8

```

US-09-943-857-4 (1-547) x US-08-379-926A-7 (1-2045)

```

Qy 89 ThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuProGlnSerGlu 108
Db 13 TCGGTCAATGGGCTCTCCGCTGATCTCCCTTTGGTACAGTCGCTACCAATGCTCTCGAG 72
Qy 109 AspCysLeuThrIleAsnValValArgProGlyThrLysAlaGlyAlaAsnLeuPro 128
Db 73 GATTGCTGAACATTGACATTGCGGTGCGGCGGACCAACCGGAGCTCGAAGCTGCCT 132
Qy 129 ValMetLeuTrpIlePheGlyGlyPheGluIleGlySerProThrIlePhePro 148
Db 133 GTGCTGCTGGATCTTTGGCGAGGCTTTGAATCTGTTCAAGGCGATGATGATGCT 192
Qy 149 AlaGlnMetValThrLysSerValLeuMetGlyLysHisIleIleHisValAlaValAsn 168
Db 193 ACAACGATGATCATCGTCGATACAGAACACATGCTTATGTTTGTAGCAATGAAT 252
Qy 169 TyrArgValAlaSerTrpGlyPheLeuAlaGlyAspIleLysAlaGluGlySerGly 188
Db 253 TATCGGTTGGGAGGTTTCGGGTTCTTGGCCGAAAGGAGATCTCGAGACGGTCCGCG 312
Qy 189 AsnAlaGlyLeuLysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGly 208
Db 313 AACCTAGGCGCTCTCGGACCAACGCTTGCCTGTCAGTGGGTTGCCGCAACATCGAGGCC 372

```

Fri Aug 6 10:53:11 2004

us-09-943-857-4.rn1

209 PheGlyGlyAspProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerVal 228
 373 TTTGGTGGACCCGGCAAGGTGACGATTTGGGAGAAATCAGCAGGACCATTCGTTG 432
 229 -LeuGlyHisLeuIleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheAr 248
 433 ACTAGATGACTTGTAC-----GACGGAACATCATTACAAGGATAAGCCCTTGTTCG 486
 248 gAlaGlyLeuMetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTyrGl 268
 487 GGGGGCCCATCATGACTCCGGTAGTGTGTTCCCGCAGACCCCTCGATGGGGTCAAGGG 546
 268 yAsnGluIleTyrAspLeuPheValSerAlaGlyCysGlySerAlaSerAspLysLe 288
 547 ACAGCAAGTATATGCGGTAGTGGATCTGCGAGGCTGTTCTTCTTAACGACACCT 606
 288 uAlaCysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGl 308
 607 AGCTTGTCTGGTGAACCTAGACTACACCGACTTCTCAATCGGCAAACTCCGTGCCAGG 666
 308 yPheLeuAlaTyrSerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIl 328
 667 CAITTTAAGCTACCATTCCTGGCGTTATCATATGCTCTGACCGGACGGACGGCGTT 726
 328 eThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleIl 348
 727 GTGGCATCACCGGACGTTTGGCAAGCAGGGAATATGCTCGGTCCGTTTCATCGT 786
 348 eGlyAspGlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrAs 368
 787 GGGCGCACCAAGAGGATGAGGAGCTTATTCGCTTGTTCAGTCCAACTTACGACGAT 846
 368 nAlaGlnAlaArgAlaTyrPheLysGlnSerPheIleHis---AlaSerAspAlaGluIl 387
 847 CGACGAGGTGGTCGACTACCTGGCTCATCTACTTCTTATGACCTAGCCGAGACGCT 906
 387 eAspThrLeuMetAlaAlaTyrProGlnAspIleThrGlnGlySerProPhe-AspThrG 407
 907 TGAAGAACTAGTGGCCCTGTACCCAGACACACCGTACGGGTCTCCGTTTCAGACAGG 966
 407 lYValLeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuA 427
 967 CGGCC--AACAACTGGTATCGCAATTTAAGCATTTGGCCGCACTTCTCGCGCATTTGG 1023
 427 lApheIleHisAlaArgArgTyrPheLeuAsnHisPheGlnGlyThr-----443
 1024 TCTTTCAC-CATTTACCGCGGGCATTCCTCTGTATGACGAGGAATCTCCCTGATCTTC 1082
 444 --LysTyrSerPheLeuSerLysGlnLeuSerGlyLeuProIleMetGlyThrPheHisA 463
 1083 CGAACTGGTCGTACCTGGGACCTATGACTATGGCACCCCACTTCTGGGGACCTTCCACG 1142
 463 lAsnAspIleValTrpGlnAspTyr-----LeuLeuGlySerGlySerV 478
 1143 GAAGTGACCTGCTGACAGGTGTTCTATGGGATCAAGCCAACTATGACGATAGTTCTAGCC 1202
 478 alIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThr---AlaG 497
 1203 ACAGGTAC-----TATCTAGCTTTGTGTATACGCTGGATCCGAATCCCAACCGGG 1253
 497 lYLeuLeuValAsnTrpProLysTyrThrSerSerSerGlnSerGlyAsnAsnLeuMetM 517
 1254 GGGAGTACATTGAGTGGCCGAGTGAAGGAATCGGGCAG-----TTGATGA 1301
 517 etIleAsnAlaLeuGlyLeuTyrThrGlyLysAsnAsnPheArgThrAlaGlyTyrAspA 537
 1302 ATTTCGGAGCGAAGCAGCGCAGTCTCTTACCGATGATTTCCGCAACCGGACATATGAGT 1361
 537 lAlaMetThrAsnProSerSerPhePheVal 547
 1362 TCATCTCTGCAGAAATACCGCGCGTTCACATC 1393

RESULT 5

US-07-846-181-6
 ; Sequence 6, Application US/07846181
 ; Patent No. 5360732
 ; GENERAL INFORMATION:
 ; APPLICANT: BERKA, RANDY M
 ; APPLICANT: FOWLER, TIMOTHY
 ; APPLICANT: REY, MICHAEL W
 ; TITLE OF INVENTION: PRODUCTION OF ASPERGILLUS NIGER
 ; TITLE OF INVENTION: CATALASE-R
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: GENECOR INTERNATIONAL, INC.
 ; STREET: 180 KIMBALL WAY
 ; CITY: SOUTH SAN FRANCISCO
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94080
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/846.181
 ; FILING DATE: 19920304
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: HORN MB, MARGARET A
 ; REGISTRATION NUMBER: 33401
 ; REFERENCE/DOCKET NUMBER: GC204-US1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-742-7536
 ; TELEFAX: 415-742-7217
 ; INFORMATION FOR SEQ ID NO: 6:
 ; LENGTH: 8533 base pairs
 ; TYPE: NUCLEIC ACID
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 US-07-846-181-6
 Alignment Scores:
 Pred. No.: 6.63e-60 Length: 8533
 Score: 632.00 Matches: 198
 Percent Similarity: 59.12% Conservative: 71
 Best Local Similarity: 43.53% Mismatches: 157
 Query Match: 22.07% Indels: 41
 DB: 1 Gaps: 10
 US-09-943-857-4 (1-547) x US-07-846-181-6 (1-8533)
 QY 108 GluAspCysLeuThrIleAsnValValArgProGlyThrLysAlaGlyAlaAsnLeu 127
 Db 12 GAGGATTGCTCAACATTGACATTCGGCGCCCGAGCAACCCCACTCGCGCAATGC 71
 QY 128 ProValMetLeuTrpIlePheGlyGlyPheGlyPheGlyLeuIleGlySerProThrIlePhePro 147
 Db 72 CCGTGTGTCTCGGATCTTTCGGGAGGCTTTGAACCTTGAAGCCCATGTATGAC 131
 QY 148 ProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIleIleHisValAlaVal 167
 Db 132 GGCACAACTGGTATCATCTGTCGATAGACACAGAATATCCCTATCGTGTGTAGCGATG 191
 QY 168 AsnTyrArgValAlaSerTrpGlyPheLeuAlaGlyAspAspIleLysAlaGluGlySer 187
 Db 192 AACTATCGGTAGGGGCTTCGGGTTCTGCGGAAAGAGAGATTCTGGAGGACGGGTCC 251
 QY 188 GlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAla 207
 Db 252 GCCAACTTAGTCTTT--GACCAGGCTTTG---CCCTAGTGGGTGCCGCAACATCGAG 306
 QY 208 GlyPheGlyGlyAspProSerLysValThrIlePheGlyGluSerAlaGlySerMetSer 227

```
Db 307 GCGTTTGGTGAGA-CGAGACAAGGTGACAATCTGGGAGAAATCAGACGGGCTATTCT 365
Qy 228 ValLeuCysHisLeuIleTyrAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPhe 247
Db 366 GTCTTGAT-CAGATGATCG-TAGACGGAACATCGCTTACAGGACGACCCCTGTGTT 423
Qy 248 ArgAlaGlyIleMetGlnSerGlyAlaMet-ValProSerAspProValAspGlyThrTy 267
Db 424 CGGA-GCCATCATGAGCTCCGGT---ATGTGTTCCCGCAGACCTGTGCGAGGGTCAA 479
Qy 267 xGlyAsn-GluIleTyrAspLeuPheValSerSerAlaGlyCysGlySerAlaSerAspL 287
Db 480 GGGATCAGCAGATATATGATCGGTGTGGACTCTGCAGGCTGTCTCTTCCACGACA 539
Qy 287 ySLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrP 307
Db 540 CCCTGGCTGTCTCGGTGAGTACACTACCGGATCTCTCAATGCGGAACCTGGT-C 598
Qy 307 roGlyPheLeuAlaTyrSerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysA 327
Db 599 CGGGATCCTAGGTATC--ACCGTGGCTATCATATGCTCGACCTCGACGAGCGGAC-G 655
Qy 327 snIleThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValI 347
Db 656 GCATTTCGCGGTGCGCAGATTTCGGTAAGCAGGGAAGTATGCGGGTCCCATCA 715
Qy 347 leIleGlyAspGlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrT 367
Db 716 TCGTGGCGACCAAGAGATGAGGGACCTGTTCGCTGTT-TCAGTC---CTTACGA 771
Qy 367 hrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPheIleHis---AlaSerAspAlaG 386
Db 772 CGATCGACGAGGTAGTCGACTATCTGGGCACCTACTTCTTCTATGACGCTAGCCAGGAC 831
Qy 386 luleAspThrLeuMetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspT 406
Db 832 AGCTTGAGAAATAGTGGCCCTGTATCCACGACACCAACATATGATGGGTCTCCCTTCAGGA 891
Qy 406 hrGlyValLeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspL 426
Db 892 CGGGCAG-GCCACAACCTGTATCCGCAATTTAAGCATTTGGCGCCCATTTCTCGGCACT 950
Qy 426 euAlaPheIleHisAlaArgArgTyrPheLeuAsnHisPheGlnGlyThr----- 443
Db 951 TGGTCTTCCATTACCCGCGGCA-TTCCTGTATATGACAGGAGCTCTCCCTCGACC 1009
Qy 444 -----LysTyrSerPheLeuSerLysGlnLeuSerGlyLeuProIleMetGlyThrPheH 462
Db 1010 TCCGAAATGGTGTACCTGGCGACCTATGACTATGGCAG-CCAATTCTGGGACCTTCC 1068
Qy 462 isAlaAsnAspIleValTyrPlnAspTyr-----LeuLeuGlySerGlyS 477
Db 1069 ATGGAAGTACCTGCTGAGGTGTTCTATGGGATCAAGCGAACTATGACGCGGATTCGA 1128
Qy 477 erValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThr---A 496
Db 1129 GCCACACGTAT-----TATCAGTTTTGTATACAGCTGGATCCGAATCCCAATC 1179
Qy 496 laGlyLeuLeuValAsnThrProLysTyrThrSerSer---SerGln-SerGlyAsnAsn 514
Db 1180 GGGGGAGGTACATGGGAATGCCCACTGGCGAGCCGACAGCTGATGATTAATTCGAGCGAAC 1239
Qy 515 LeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPheGthrAlaGly 534
Db 1240 AGCAGTCTCTTACG-----GATATTTCGCAACGGGACA 1275
Qy 535 TyrAspAlaLeuMetThrAsnProSerSerPhePheVal 547
Db 1276 TATGAGTTTATCTCGAGAATACCGCGGCTTCCACATC 1314
```

RESULT 6
US-07-845-989-6

```
; Sequence 6, Application US/07845989
; Patent No. 5360901
; GENERAL INFORMATION:
; APPLICANT: BERKA, RANDY M
; APPLICANT: FOWLER, TIMOTHY
; APPLICANT: REY, MICHAEL W
; TITLE OF INVENTION: PRODUCTION OF ASPERGILLUS NIGER
; TITLE OF INVENTION: CATALASE-R
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENENCOR INTERNATIONAL, INC.
; STREET: 180 KIMBALL WAY
; CITY: SOUTH SAN FRANCISCO
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/845,989
; FILING DATE: 19920304
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: HORN Ms, MARGARET A
; REGISTRATION NUMBER: 33401
; REFERENCE/DOCKET NUMBER: GC208-US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-742-7536
; TELEFAX: 415-742-7217
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8533 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-07-845-989-6

Alignment Scores:
Pred. No.: 6.63e-60 Length: 8533
Score: 632.00 Matches: 198
Percent Similarity: 59.12% Conservative: 71
Best Local Similarity: 43.52% Mismatches: 157
Query Match: 22.07% Indels: 41
DB: 1 Gaps: 10

US-09-943-857-4 (1-547) x US-07-845-989-6 (1-8533)

Qy 108 GluAspCysLeuThrIleAsnValValArgProGlyThrLysAlaGlyAlaAsnLeu 127
Db 12 GAGGATTTCTGAACATTTGACATTCGGCGCGCCAGCGAAACCCCACTGCGGCGAATGC 71
Qy 128 ProValMetLeuTrpIlePheGlyGlyPheGluIleGlySerProThrIlePhePro 147
Db 72 CCGTGTGCTCTCGGATCTTTGGCGAGGCTTTGAACCTTGGTTCAAAGGCCATGTATGAC 131
Qy 148 ProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIleIleHisValAlaVal 167
Db 132 GGCACAAACGATCGTATCATCGTAGACAGAAATATCCCTATCGTGTGTAGCGATG 191
Qy 168 AsnTyrArgValAlaSerTrpGlyPheLeuAlaGlyAspIleLysAlaGlyLysSer 187
Db 192 AACTATCGGTAGGGGCTTCGGGTTCTGCGCGGAAAGAGATTCTGAGGACGGGTCC 251
Qy 188 GlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAla 207
Db 252 GCCAACTTAGGTCTTT--GACCAAGCCTTG---CCCTAGTGGGTGCCCAACATCGAG 306
Qy 208 GlyPheGlyGlyAspProSerLysValThrIlePheGlyGluSerAlaGlySerMetSer 227
```


Db 239 ACGCACAGGCGCTCGCCAGCGGACTGAGTATGAGACCGATCTGTGGCCCGCTAGC--- 295
Qy 79 GluGlyThrPheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLys 98
Db 296 GAGGGAACC-----AACAGCAGCAGGCGCTGCGGCCCTTAGCAAC 337
Qy 99 ValPheGlnAlaValLeuProGln-----SerGluAspCysLeuThrIleAsnVal 115
Db 338 AGCTCGAGCAGCGG---CCGAGAAACAGCGCTCGGAGGATGCTCTTCCTCAATGTC 394
Qy 116 ValArgProGlyThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGly 135
Db 395 GTTCCCGCCCGCGCTCGTGGCAGGCGACAACTCTCCGCTCGCTACATCAGCGA 454
Qy 136 GlyGlyPheGluIleGlySerProThrIlePheProAlaGlnMetValThrLysSer 155
Db 455 GGTGGCTACGCTTCGGCGATGCGAGCAGCGGACGACCTTTCGCGCTTCACCAAGCAC 514
Qy 156 ValLeuMetGlyLysHisIleHisValAlaValAsnTyrArgValAlaSerTrpGly 175
Db 515 ACG-----GGAACCAAGATGGTGTGTAATCTCCAGTACCGCTCTCGGCGCTTGGT 568
Qy 176 PheLeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGln 195
Db 569 TTCTCTGCTGCCAGCCATGAGGACTACGCTGTAACGACGCGGCTGCTTGACACAG 628
Qy 196 ArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLys 215
Db 629 CAATTCGCCCTTCAATGGGTTCAACAGCAGCTCTCGAAGTTCGGCGGCAACCCGATCAC 688
Qy 216 ValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsn 235
Db 689 GTTACGATTTGGGCGGAGTCTGAGCGCGAGGTCCTGTTATGAACACGATCAITGCGAAC 748
Qy 236 AspGlyAspAsnThrTyrLysGly-----LysProLeuPheArgAlaGlyIle 251
Db 749 ---GGCGGCAACCGCTCAAGGCTCTCGTCTCAAGAGCCCTCTTCCACGCTGCCATC 805
Qy 252 MetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGluIle 271
Db 806 GGCTCTCTCGCTTCCTCCCTCCACCAAGCAGTACAACTCCCTTCGCGGAGTCTC 865
Qy 272 TyrAspLeuPheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCysLeu 291
Db 866 TACTCCCAACTCTCTCGGCGACAACTGCACCAAGCCGCTCGCTTCGCTCGCTC 925
Qy 292 ArgSerAlaSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeuAla 311
Db 926 GAAGCTGTCAGCGCTCGGCGCTC-----GCTCGCGCGGCGGTGAAGAACTCGCGCGCG 979
Qy 312 TyrSerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThrAspAsp 331
Db 980 TTCCCGTTCGGGTTTGGTGTATGTCCTCGGTCGTCAGCGGACCTTCGTGATGAGCGC 1039
Qy 332 MetTyrLysLeuValArgAspGly---LysTyrAlaSerValProValIleIleGlyAsp 350
Db 1040 GGTGCTGCTCTCTCCCAAGGCGAAGAACTCAATGGCAACCTCTTCCCGGATC 1099
Qy 351 GlnAsn---AspGluGlyThrIlePheGlyLeuSerSerLeu---AsnValThrThrAsn 368
Db 1100 ACAACCTCGAGAGGATTCATTTACTGAGCCACCTATTGAGAACGACGATCAGC 1159
Qy 369 AlaGlnAlaArgAlaTyrPheLysGlnSerPheIleHisAlaSerAspAlaGluIleAsp 388
Db 1160 GACCACTGCGAGCGCTC-----TCCAGTTCGAC 1189
Qy 389 ThrLeuMetAlaAla---TyrPro-----GlnAspIleThrGln 400
Db 1190 CGCTCTCTCGCGGCTCTTCCCTATCATCATCCTCGAGGAGCGCGCGGCGGGAAG 1249
Qy 401 GlySerProPheAspThrGlyValLeuAsnAlaLeuThrProGlnPheLysArgIleSer 420
Db 1250 CAGTACCCGATCTCCAGCGCGCGCTCAAGGGCAACACC-----TTCTCTCGCATCTCG 1303

Qy 421 AlaValLeuGlyAspLeuAlaPheIleHisAlaArgTyrPheLeuAsnHisPheGln 440
Db 1304 GCGTATCGCGGACTCGACCTTCGTC---TGCCCGACCTACTGG----- 1345
Qy 441 GlyGlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeu---ProIleMetGly 459
Db 1346 -----ACCGCGGAGGCTTCGGCTCGTCCGCCCAAGGCGCTCTTCGACTACGCGCGC 1399
Qy 460 ThrPheHisAlaAsnAspIleValTrpGlnAspTyrLeuLeuGlySer----- 475
Db 1400 GCTCACACCGACCGAC-----AACTCGTACTACATCGGCTCCATCTGGAACGGC 1450
Qy 476 -----GlySerVal---IleTyrAsnAsnAlaPheIleAlaPheAlaThrAsp 490
Db 1451 AAGAGTCTGCTCTCGTCCGCTCAGTCTTCAGCGCGCGCTCGCGGCTTCATCGAGACG 1510
Qy 491 LeuAspPro-----AsnThrAlaGlyLeuLeuValAsn-----TrpProLysTyrThr 506
Db 1511 TTCAACCGCAACAACACGCTGCCCAACAGACCATCAACCTTACTGCGCGACGCTTCGAC 1570
Qy 507 SerSerSerGln 510
Db 1571 TCGGCGCAAGCAG 1582

RESULT 8
US-07-732-962A-1
; Sequence 1, Application US/07732962A
; Patent No. 5248604
; GENERAL INFORMATION:
; APPLICANT: Fischer, Meir
; TITLE OF INVENTION: EXPRESSION OF ENZYMATICALLY ACTIVE
; TITLE OF INVENTION: RECOMBINANT HUMAN ACETYLCHOLINESTERASE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White, Esq.
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/732,962A
; FILING DATE: 19910722
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 39304/JPW/LSW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1845 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1842
; US-07-732-962A-1

Alignment Scores:
Pred. No.: 1.02e-40 Length: 1845
Score: 452.00 Matches: 156

1006 GGAGACTTCTCAGTCAGACACCCAGAGGCCCTCATCAACCGCGGAGACTTCCACGGCGCTG 1065

345 ProValIleIleGlyAspGlnAsnAspGluGlyThrIlePhe----- 358

1066 CAGGTCTGTGGTGTGGTGAAGAGTAGGGCTCGTATTTCTTCTGGTTTACGGGGCCCCA 1125

359 GlyLeuSerSerLeuAsn-----ValThrThrAsnAlaGlnAlaAargAlaTyrPheLys 376

1126 GGCTTCAGCAAAAGACACAGAGTCTCTCATCAGCGCGGGCCGAGTTCCTGGCGGGGTGCGG 1185

377 GlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyr----- 394

1186 GTCGGGGTTCCTCCAGAGTAAAGTAGCTCGCAGCGCGAGGCTGTGTCTCTGCAATTACACAGAC 1245

395 -----ProGlnAsp-----IleThrClnGlySerProPheAspThrGlyVal 408

1246 TGGCTGCATCCCGAGGAGCCGCGAGCTCAGGGAGCCCTGAGCGATGTGTGGGGCGAC 1305

409 LeuAsnAlaLeuThrProGlnPheLysAArgIleSerAlaValLeuGlyAspLeuAlaPhe 428

1306 CACAATCTGTGTGCCCC-----GTGCCCCAGCTGGGTGGCGCACTGGCTGCC 1353

429 IleHisAlaArg-----ArgTyrPheLeuAsnHisPheGlnGly----- 441

1354 CAGGGTCCCGGGGTCTACGGCTACGTCTTTGAACACCGTGCTTCCACGCTCTCTCTGGCCC 1413

442 -----GlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeuProIle--- 457

1414 CTGTGGATGGGGTGGCCCGCGCTACGAGATCGAGTTCATCTTTGGATCCCCCTGGAC 1473

458 ---MetGlyThrPheHisAlaAsnAspIleValTrpGlnAspTyrLeuGlySerGly 476

1474 CCCTCTCGAAACTACACGGCAGAGAGAAATCTTCGCCACGACACTGATG----- 1524

477 SerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAla 496

1525 -----CGATACTGGGCCAACTTTGCCCGCACAGGGATCCCAATCAGGCC 1566

497 -----GlyLeuValAsnTrpProLysTyrThrSerSerSerGln 510

1570 CGAGACCCCAAGGCCCCACCAATGCCCGGTACACGGCGGGGTCTCAG 1617

RESULT 9

PCT-US92-06106-1

; Sequence 1, Application PC/TUS9206106

; GENERAL INFORMATION:

; APPLICANT: Fischer, Meir

; TITLE OF INVENTION: ENZYMATICALLY ACTIVE RECOMBINANT HUMAN

; TITLE OF INVENTION: ACETYLCHOLINESTERASE AND USES THEREOF

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESS: John P. White, Esq.

; STREET: 30 Rockefeller Plaza

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10112

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US92/06106

; FILING DATE: 19920722

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: White, John P.

; REGISTRATION NUMBER: 28,678

; REFERENCE/DOCKET NUMBER: 39304-B-PCT/JPW/EAB

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 977-9550

; TELEFAX: (212) 664-0525

TELEPHONE: (212) 664-0100
TELEFAX: (212) 664-0101

Fri Aug 6 10:53:11 2004

CITY: Farmington Hills
STATE: Michigan
COUNTRY: US
ZIP: 48334
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/318,826A
FILING DATE:

CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Kohn, Kenneth I.
REGISTRATION NUMBER: 30,955
REFERENCE/DOCKET NUMBER: 2391.00001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (248) 539-5050
TELEFAX: (248) 539-5055

INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2256 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO

ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens

FEATURE:
OTHER INFORMATION: /note= "Splice variant: Exons 1, 2,
OTHER INFORMATION: 3, 4 and 6"
US-08-318-826A-5

Alignment Scores:
Pred. No.: 1,438-40 Length: 2256
Score: 452.00 Matches: 156
Percent Similarity: 44.53% Conservative: 83
Best Local Similarity: 29.10% Mismatches: 179
Query Match: 15.78% Indels: 118
Gaps: 26

US-09-943-857-4 (1-547) x US-08-318-826A-5 (1-2256)

QY 35 AlaPheLeuGlyIleProPheAlaGluProValGlyAsnLeuArgPheLysAspPro 54
Db 343 GTTTTCTGGGATCCCTTTGGGAGCCACCCATGGGACCCCGCTGCTTTCTGCCACCG 402
QY 55 -----ValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyPro 71
Db 403 GAGCCCAACGACGCTTGGTCAGGGGTGGTAGAGCTCAACCTTCAGAGTGTC----- 456
QY 72 SerCysMetGln-----GlnAsnProGluGlyThrPheGluGluAsn 85
Db 457 ----TGTAACAATATGTGACACCCCTATACCCAGGTTTTCAGGGCCACCGAGATGTGAAC 513
QY 86 LeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuPro 105
Db 514 CCCAACCGTGAGCTG-----GAGCTGTACGATGGCCGCTTCTTTGGTACAGGCC 675
QY 106 GlnSerGluAspCysLeuThrIleAsnValValArgProGlyThrLysAlaGlyAla 125
Db 529 ---AGCGAGGACTGCTGTACCTCAAC---GTGTGACACCATACCCCGGCTACATCC 582
QY 126 AsnLeuProValMetLeuThrIlePheGlyGlyPheGluLeuLeuSerProThrIle 145
Db 583 CCACCCCTGCTGCTGTGATCTATGGGGGTGGCTTCTACAGTGGGGCTCCTCCTTG 642
QY 146 PheProProGlnMetValThrLysSerValLeuMetGlyLysHisIleIleHis--- 164
Db 643 -----GAGCTGTACGATGGCCGCTTCTTTGGTACAGGCC 675

QY 165 -----ValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGly 179
Db 676 GAGAGGACTGCTGCTGGTGTGCATGAACTACCGGGTGGAGCCTTTGCTTCTCGCCCTG 735
QY 180 AspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMet 199
Db 736 CCGGGGAGCGA---GAGGCCCGGGCAATGTGGGTCTCTCTGGATCAGAGCTGCGCCCTG 792
QY 200 GlnTrpValAlaAspAsnIleAlaGlyPheGlyAspProSerLysValThrIlePhe 219
Db 793 CAGTGGGTGACGAGAACGTGGCAGCCTTCGGGGGTGACCGACATCAGTCAAGCTGTT 852
QY 220 GlyGluSerAlaGlySerMetSerValLeuCysHisIleLeuTrpAsnAspGlyAspAsn 239
Db 853 GGGGAGAGCGCGGAGCGCTCGTGGTGGCATGACCTGCTG----- 894
QY 240 ThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValPro 259
Db 895 TCCCCCGCCAGCCGGGCTTGTCCACAGGCCGTGTGTCAGAGCGGTGCC-----CCC 948
QY 260 SerAspProValAspGlyThrTyrGly-----AsnGluIle 271
Db 949 AATGGACCTGG---GCCACGCTGGCATGGAGAGGCCCTGCGCAGGCGCCACGAGCTG 1005
QY 272 TyrAspLeuPheValSerSerAlaGlyCys-----GlySerAlaSer 285
Db 1006 GCCCACCTT-----GTGGCTGTCTCTCAGGCGGACCTGTTGGGAATACACACA 1053
QY 286 AspLysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsn 305
Db 1054 GAGCTGTAGCTGCTTCGGACACGACCCAGCGAGGTCTCTGGTG-----AACCA 1104
QY 306 ThrProGlyPheLeuAlaTyrSerSerLeu---ArgLeuSerTyrLeuProArgPro 324
Db 1105 GAATGGCAGCTGCTGCTCAAGAAAGCTTCTCCGGTTCTCTCTGCTGTGTAGTAG 1164
QY 325 GlyLysAsnIleThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerVal 344
Db 1165 GGAGACTTCTCTCAGTGACACCCAGAGGCCCTCATCAACGGCGGAGACTTCCACGGCCCTG 1224
QY 345 ProValIleIleGlyAspGlnAsnAspGlyThrIlePhe----- 358
Db 1225 CAGGTGCTGGTGGTGGTGAAGATGAGGCTCGTATTCTGTTTACGGGGCCCA 1284
QY 359 GlyLeuSerSerLeuAsn-----ValThrThrAsnAlaGlnAlaArgAlaTyrPheLys 376
Db 1285 GGCTTCAGCAAGACACACGAGTCTCTCATCAGCCGGCCGAGTCTCTGCGCGGGTGGCG 1344
QY 377 GlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyr----- 394
Db 1345 GTCGGGTTTCCCAAGTAACTGACCTGGCAGCGGAGGTGTGCTCTGCTTACACAGAC 1404
QY 395 -----ProGlnAsp-----IleThrGlnGlySerProPheAspThrGlyVal 408
Db 1405 TGCGTGCATCCCGAGGACCCCGACCCCTGAGGGAGCCCTGAGCGATGTGGTGGCGAC 1464
QY 409 LeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPhe 428
Db 1465 CACAAATGCTGTGGCCCC-----GTGGCCAGCTGGCTGGGCGACTGGCTGCC 1512
QY 429 IleHisAlaArg-----ArgTyrPheLeuAsnHisPheGlnGly----- 441
Db 1513 CAGGGTGGCCGGTCTACGCTTGTAAACACCGTGTCTCCAGCTCTCCTCGGCC 1572
QY 442 -----GlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeuProIle--- 457
Db 1573 CTGTGATGGGGTGGCCACCGCTACGAGATCGATTCATCTTTGGATCCCTCGGAC 1632
QY 458 ---MetGlyThrPheHisAlaAsnAspIleValTrpGlnAspTyrTrpLeuGlySerGly 476
Db 1633 CCCTCTCGAAACTACACGGCAGAGGAGAAATCTTGGCCAGCGACTGATG----- 1683

QY 477 SerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAla 496
 Db 1684 -----CGACTATGGGCCCAACTTTTGGCCGACACAGGGGATCCCAATGAGCCC 1728
 QY 497 -----GlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGln 510
 Db 1729 CGAGACCCCAAGGCCACATGCCCCCGTACAGCGCGGGGCTCAG 1776

RESULT 11
 US-08-370-156-1
 ; Sequence 1, Application US/08370156
 ; Patent No. 5932780
 ; GENERAL INFORMATION:
 ; APPLICANT: Soreq, Hermona
 ; APPLICANT: Zakut, Haim
 ; APPLICANT: Shani, Moshe
 ; TITLE OF INVENTION: TRANSGENIC ANIMAL ASSAY SYSTEM FOR
 ; TITLE OF INVENTION: ANTICHOLINESTERASE SUBSTANCES
 ; NUMBER OF SEQUENCES: 27
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Reising, Ethington, Barnard & Perry
 ; STREET: P. O. Box 4390
 ; CITY: Troy
 ; STATE: Michigan
 ; COUNTRY: US
 ; ZIP: 48099
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/370,156
 ; FILING DATE:
 ; CLASSIFICATION: 536
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Kohn, Kenneth I.
 ; REGISTRATION NUMBER: 30,955
 ; REFERENCE/DOCKET NUMBER: P-307 (Mulford)
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (810) 689-3500
 ; TELEFAX: (810) 689-4071
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2256 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-370-156-1

Alignment Scores:
 Pred. No.: 1.43e-40 Length: 2256
 Score: 452.00 Matches: 156
 Percent Similarity: 44.59% Conservative: 83
 Best Local Similarity: 29.10% Mismatches: 179
 Query Match: 15.78% Indels: 118
 DB: 2 Gaps: 26

US-09-943-857-4 (1-547) x US-08-370-156-1 (1-2256)

QY 35 AlaPheLeuGlyIleProPheAlaGluProValGlyAsnLeuArgPheLysAspPro 54
 Db 343 GCTTTCTGGGATCCCTTTGGGAGGCCACCCATGGGAGCCCGCTTCTGCGCACCG 402
 QY 55 -----ValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyPro 71
 Db 403 GAGCCCAAGCAGCCTTGGTCAGGGGTGGTAGCGTACAACTTCCAGAGTGC----- 456
 QY 72 SerCysMetGln-----GlnAsnProGluGlyThrPheGluGluAsn 85
 Db 457 ---TGCTACCAATATGTGGACACCCCTATACCCAGGTTTGGAGGGACCGAGATGTGGAC 513
 QY 86 LeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuPro 105

Db 514 CCCAACCGTGAGCTG----- 528
 QY 106 GlnSerGluAspCysLeuThrIleAsnValValArgProProGlyThrLysAlaGlyAla 125
 Db 529 ---AGCAGGAGCTGCTGTACCTCAAC---GTGTGGACACCATACCCCGCCCTACATCC 582
 QY 126 AsnLeuProValMetLeuTyrIlePheGlyGlyGlyPheGluIleGlySerProThrIle 145
 Db 583 CCCACCCCTGCTCGTCTGGATCTATGGGGTGGCTTCTACGTGGGGCTCTCTCTCTG 642
 QY 146 PheProProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIleIleHis--- 164
 Db 643 -----GACGTGTACGATGCCCGCTTCTGGTACAGGCC 675
 QY 165 -----ValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGly 179
 Db 676 GAGAGGACTGTGCTGGTGTCTCATGAACACCGGGTGGAGCCTTTGGCTTCTGGCCCTG 735
 QY 180 AspAspIleLysAlaGluGlySerGlyAsnAlaGlyLysAspGlnArgLeuGlyMet 199
 Db 736 CCGGGAGCCGA---GAGGCCCGGCAATGTGGTCTCTGTGATCAGAGCTGGCCCTG 792
 QY 200 GlnTrpValAlaAspAsnIleAlaGlyPheGlyAspProSerLysValThrIlePhe 219
 Db 793 CAGTGGGTGCAGGAGAACGTGGCAGCCTTCGGGGGTGACCCGACATCAGTACGCTGTT 852
 QY 220 GlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAspGlyAspAsn 239
 Db 853 GGGGAGAGCGCGGAGCGCTCGTGGGATGCACCTGCTG----- 894
 QY 240 ThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValPro 259
 Db 895 TCCCGCCCGAGCGGGCCTGTTCACAGGGCGCTGCTGCAGAGCGGTGCC-----CCC 948
 QY 260 SerAspProValAspGlyThrTyrGly-----AsnGluIle 271
 Db 949 AATGAGCCCTGG---GCCAGGTGGCATGGAGAGGCCCGTCGAGGGCCAGCAGCTG 1005
 QY 272 TyrAspLeuPheValSerSerAlaGlyCys-----GlySerAlaSer 285
 Db 1006 GCCCACCTT-----GTGGGTGTCTCCAGCGGCCTGTGTGGGAATGACACA 1053
 QY 286 AspLysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsn 305
 Db 1054 GAGCTGTAGCTTGCCTTCGACACACGACCGCAGGCTCTGTGTG-----AACAC 1104
 QY 306 ThrProGlyPheLeuAlaTyrSerSerLeu---ArgLeuSerTyrLeuProArgProAsp 324
 Db 1105 GAATGGCAGCTGCTGCCTCAAGAAAGGCTTCTCGGTCTCTCTGCTGCTGTGGTAGAT 1164
 QY 325 GlyLysAsnIleThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerVal 344
 Db 1165 GGAGACTTCTCAGTGACACCCCGAGAGCCCTCATCAACGCGGAGACTTCCACGCGCTG 1224
 QY 345 ProValIleIleGlyAspGlnAsnAspGluGlyThrIlePhe----- 358
 Db 1225 CAGTGTCTGGTGTGTGTGAAGATGAGGCTCGTATTTTCTGTTTACGGGGGCCCA 1284
 QY 359 GlyLeuSerSerLeuAsn-----ValThrThrAsnAlaGlnAlaArgAlaTyrPheLys 376
 Db 1285 GGCTTCAGCAAGACACACAGCTCTCTCATCAGCCCGCGCCGAGTTCCTGCGCGGGGTGCGG 1344
 QY 377 GlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyr----- 394
 Db 1345 GTCGGGTTTCCCGAGGTAACTGACCTGGCAGCGGAGGTGTGTCTCTGTCATATACAGAC 1404
 QY 395 -----ProGlnAsp-----IleThrGlnGlySerProPheAspThrGlyVal 408
 Db 1405 TGGCTGTCATCCCGAGACCCCGCAGCCCTGAGGGAGGCCCTGAGCGATGTGGTGGCGAC 1464
 QY 409 LeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPhe 428

Db 1465 CACAACTGCTGTGCCCC-----GTGGCCCGAGCTGGCTGGCGAGCTGGCTGCC 1512
 Qy 429 IleHisAlaArg-----ArgTyrPheLeuAsnHisPheGlnGly----- 441
 Db 1513 CAGGGTGGCCGGGTCTACGCTAGCTCTTTTGAACACCGCTGCTCCACGCTCTCTCTGGGCC 1572
 Qy 442 -----GlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeuProile--- 457
 Db 1573 CTGTGGATGGGGGCCCGCTACGAGATCGAGTTCATCTTTGGGATCCCTGGAC 1632
 Qy 458 ---MetGlyThrPheHisAlaAsnAspIleValTyrGlnAspTyrLeuLeuGlySerGly 476
 Db 1633 CCCTCTCGAACTACACGGCAGAGGAGAAATCTTGGCCCGAGCTGATG----- 1683
 Qy 477 SerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAla 496
 Db 1684 -----CGATACTGGGCCCACTTTGCCCGCACAGGGGATCCCAATGAGGCC 1728
 Qy 497 -----GlyLeuLeuValAsnTyrProLysTyrThrSerSerSerGln 510
 Db 1729 CGAGACCCCAAGGCCCAATGGCCCCCGTACAGCGGGGGCTCAG 1776

RESULT 12

US-08-814-095-1
 ; Sequence 1, Application US/08814095
 ; Patent No. 6025183
 ; GENERAL INFORMATION:
 ; APPLICANT: Soreq, Hermona
 ; APPLICANT: Zakut, Haim
 ; APPLICANT: Shani, Moshe
 ; TITLE OF INVENTION: TRANSGENIC ANIMAL ASSAY SYSTEM FOR
 ; TITLE OF INVENTION: ANTI-CHOLINESTERASE SUBSTANCES
 ; NUMBER OF SEQUENCES: 7
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: KOHN & ASSOCIATES
 ; STREET: 30500 No. 6025183thwestern Highway, Suite 410
 ; CITY: Farmington Hills
 ; STATE: Michigan
 ; COUNTRY: U.S.
 ; ZIP: 48334
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/814,095
 ; FILING DATE:
 ; CLASSIFICATION: 800
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Montgomery, Ilene N.
 ; REGISTRATION NUMBER: 38,972
 ; REFERENCE/DOCKET NUMBER: 2391.00066
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (248) 539-5050
 ; TELEFAX: (248) 539-5055
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2256 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: other nucleic acid
 ; DESCRIPTION: /desc = "ACHE gene comprising exons
 ; DESCRIPTION: 2, 3, 4 and 6"
 ; HYPOTHEICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Homo sapiens
 ; US-08-814-095-1

Alignment Scores: 1.43e-40 Length: 2256
 Pred. No.:

Score: 452.00 Matches: 156
 Percent Similarity: 44.59% Conservative: 83
 Best Local Similarity: 29.10% Mismatches: 179
 Query Match: 15.78% Indels: 118
 DB: 3 Gaps: 26
 US-09-943-857-4 (1-547) x US-08-814-095-1 (1-2256)
 Qy 35 AlaPheLeuGlyIleProPheAlaGluProValGlyAsnLeuArgPheLysAspPro 54
 Db 343 GCTTTCCTGGCATCCCTTTGGGAGCCACCCATGGAGCCCGCTCGCTTCTGCGACCG 402
 Qy 55 -----ValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyPro 71
 Db 403 GAGCCCAAGCAGCCTTGTGTCAGGGGTGTAGACCTTCCAGAGTGC----- 456
 Qy 72 SerCysMetGln-----GlnAsnProGluGlyThrPheGluGluAsn 85
 Db 457 ---TGCTACCAATATGGGACACCTATACCCAGGTTTGGAGCCACCGAGATGTGGAC 513
 Qy 86 LeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuPro 105
 Db 514 CCCAACCGTGAGCTG----- 528
 Qy 106 GlnSerGluAspCysLeuThrIleAsnValValArgProGlyThrLysAlaGlyAla 125
 Db 529 ---AGCGAGGACTGCTGTACTCAAC---GTGTGGACACCATACCCCGGCTCATATCC 582
 Qy 126 AsnLeuProValMetLeuTyrIlePheGlyGlyPheGluIleGlySerProThrIle 145
 Db 583 CCCACCCTGTCTCTGTGATCTATGGGGTGGCTTCTACAGTGGGGCTCTCTCTG 642
 Qy 146 PheProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIleIleHis--- 164
 Db 643 -----GAGGTGTACGATGGCGCTTCTGTGTACAGGCC 675
 Qy 165 -----ValalaValAsnTyrArgValAlaSerTyrPheLeuAlaGly 179
 Db 676 GAGGAGACTGTCTGTCTCCATGAACACCGGGTGGGAGCCTTTGGCTTCTGCGCCCTG 735
 Qy 180 AspAspIleLysAlaGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMet 199
 Db 736 CCGGGAGCCGA---GAGCCCGGGCAATGTGGGTCTCTCTGATCAGAGGTGGCCCTG 792
 Qy 200 GlnTyrValAlaAspAsnIleAlaGlyPheGlyAspProSerLysValThrIlePhe 219
 Db 793 CAGTGGTGCAGGAGACGTGGCAGCTTCGGGGGTGACCCGACATCAGTACGCTGTTT 852
 Qy 220 GlyLysSerAlaGlySerMetSerValLeuCysHisLeuIleTyrAsnAspGlyAspAsn 239
 Db 853 GGGGAGAGCGCGGAGCGCCTCGTGGGCATGCACTGCTG----- 894
 Qy 240 ThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValPro 259
 Db 895 TCCCGCCCGAGCGGGGCTGTTCCACAGGGCGGTGCTGCAGAGCGGTGCC-----CCC 948
 Qy 260 SerAspProValAspGlyThrTyrGly-----AsnGluIle 271
 Db 949 AATGGACCTGG---GCCACGGTGGGATGGAGAGCCCGTGGAGGCCACGACGCTG 1005
 Qy 272 TyrAspLeuPheValSerSerAlaGlyCys-----GlySerAlaSer 285
 Db 1006 GCCCACCTT-----GTGGGCTGTCTCCAGGGCGGCACTGTGGGATGACACA 1053
 Qy 286 AspLysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsn 305
 Db 1054 GAGCTGTAGCTGCTTGGACACCGACCGAGCGAGCTCTGCTG-----AACAC 1104
 Qy 306 ThrProGlyPheLeuAlaTyrSerSerLeu---ArgLeuSerTyrLeuProArgProAsp 324
 Db 1105 GAATGGCAGCTGTGCTCTCAAGAAAGCGTCTTCCGGTCTCTCTGCTGTGGTGTAGAT 1164
 Qy 325 GlyLysAsnIleThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerVal 344


```
Db 949 AATGACACCTGG---GCCACGGTGGGATGGAGAGGCCCGTCGACGGGCCACCGCAGCTG 1005
Qy 272 TyrAspLeuPheValSerSerAlaGlyCys-----GlySerAlaSer 285
Db 1006 GCCACACCTT-----GTGGCTGTCTCCACGGCGGCACCTGTGGGAATGACACA 1053
Qy 286 AspLysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeuAlaThrAsnAsn 305
Db 1054 GAGTGTGTAGCTGCTTTCGACACACGACGACGAGCTCTCTGGTG-----AACCAC 1104
Qy 306 ThrProGlyPheLeuAlaTyrSerSerLeu---ArgLeuSerTyrLeuProArgProAsp 324
Db 1105 GAATGGCAGCTGCTCCCTCAAGAAAGCTCTCCGTTCTCTCTCGTCCCTGTGTAGAT 1164
Qy 325 GlyLysAsnLeuThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerVal 344
Db 1165 GGAGACTTCTCTCAGTGACACCCACGAGGCCCTCATCAACGGGGAGACTTCCACGGCCTG 1224
Qy 345 ProValIleLeuGlyAspGlnAsnAspGluGlyThrIlePhe----- 358
Db 1225 CAGGTGTCTGGTGGTGTGTGAAGATCAGGCTCGTATTTCTGTTTACGGGGCCCA 1284
Qy 359 GlyLeuSerSerLeuAsn-----ValThrThrAsnAlaGlnAlaArgAlaTyrPheLys 376
Db 1285 GGCTTCAGCAAGACAACAGCTCTCTCATCGCGGGCGAGTTCCTGGCGGGGTGCGG 1344
Qy 377 GlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyr----- 394
Db 1345 GTCGGGTTCCTCCAGGTAAAGTGACCTGCAGCCGAGGCTGTGGTCTCTGATACACAGAC 1404
Qy 395 -----ProGlnAsp-----IleThrGlnGlySerProPheAspThrGlyVal 408
Db 1405 TGGCTGCATCCGAGGACCCGCGACGCTGAGGAGGCCCTGAGCGATGTGTGGCGGAC 1464
Qy 409 LeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPhe 428
Db 1465 CACATGTGTGTGCCCC-----GTGCGCCAGCTGGCTGGCGGACTGGCTGCC 1512
Qy 429 IleHisAlaArg-----ArgTyrPheLeuAsnHisPheGlnGly----- 441
Db 1513 CAGGGTCCCGGGTCTACGGCTAGCTTTTGACACCGCTGCTCCACGCTCTCTGCGCC 1572
Qy 442 -----GlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeuProIle--- 457
Db 1573 CTGTGGATGGGGTGGCCCGCTACGAGATCGAGTTCATCTTTGGGATCCCGCTGGAC 1632
Qy 458 ---MetGlyThrPheHisAlaAsnAspIleValTyrPheGlnAspTyrLeuLeuGlySerGly 476
Db 1633 CCCTCTCGAACTACACGGCAGAGGAGAAATCTTCGCCCGACGAGCTGATG----- 1683
Qy 477 SerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAla 496
Db 1684 -----CGATACTGGGCCAACTTTGGCCCGCACAGGGGATCCCAATGAGCCC 1728
Qy 497 -----GlyLeuLeuValAsnTyrProLysTyrThrSerSerSerGln 510
Db 1729 CGAGACCCCAAGGCCCCCAATAGCCCGCTACACGGCGGGGCTCAG 1776

RESULT 15
US-08-370-156-5
; Sequence 5, Application US/08370156
; Patent No. 5932780
; GENERAL INFORMATION:
; APPLICANT: Soreq, Hermona
; APPLICANT: Zakut, Haim
; APPLICANT: Shani, Moshe
; TITLE OF INVENTION: TRANSGENIC ANIMAL ASSAY SYSTEM FOR
; TITLE OF INVENTION: ANTICHOLINESTERASE SUBSTANCES
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Reising, Ethington, Barnard & Perry
; STREET: P.O. Box 4390
; CITY: Troy
```

```
; STATE: Michigan
; COUNTRY: US
; ZIP: 48099
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/370,156
; FILING DATE:
```

```
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,955
; REFERENCE/DOCKET NUMBER: P-307 (Mulford)
; TELEPHONE: (810) 689-3500
; TELEFAX: (810) 689-4071
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3016 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 160..2010
; US-08-370-156-5
```

```
Alignment Scores:
Pred. No.: 2,32e-40 Length: 3016
Score: 452.00 Matches: 156
Percent Similarity: 44.59% Conservative: 83
Best Local Similarity: 29.10% Mismatches: 179
Query Match: 15.78% Indels: 118
DB: 2 Gaps: 26
```

```
US-09-943-857-4 (1-547) x US-08-370-156-5 (1-3016)
Qy 35 AlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArgPheLysAspPro 54
Db 343 GCTTTCCTGGCATCCCTTTGGGAGCCACCCATGGGACCCCGTCGCTTCTGCCACCG 402
Qy 55 -----ValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyPro 71
Db 403 GAGCCCAAGCAGCCTTGGTCAGGGGTGTAGACGCTACACCTTCCAGAGTGTC----- 456
Qy 72 SerCysMetGln-----GlnAsnProGluGlyThrPheGluGluAsn 85
Db 457 ---TGCTACCAATATGTGGACACCTATACCCAGGTTTGGAGGCGACCGAGATGTGAAC 513
Qy 86 LeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuPro 105
Db 514 CCACACCGTGAGCTG-----GlnAsnProGluGlyThrPheGluGluAsn 125
Qy 106 GlnSerGluAspCysLeuThrIleAsnValValArgProProGlyThrLysAlaGlyAla 125
Db 529 ---AGGAGAGACTGCTGTACCTCAAC---GTGTGGACACCATACCCCGGCTCATATCC 582
Qy 126 AsnLeuProValMetLeuTyrIlePheGlyGlyGlyPheGluIleGlySerProThrIle 145
Db 583 CCCACCCCTGCTCGTCTGGATCTATGGGGTGGCTTCTACAGTGGGGCCTCCTCCTTG 642
Qy 146 PheProProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIleIleHis--- 164
Db 643 -----GACGTGTACGATGGCGCTTCTTGGTACAGGCC 675
Qy 165 -----ValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGly 179
Db 676 GAGAGACTGTGCTGTGTCCATGAACCTACCGGGTGGGAGCCTTTGGCTTCTCGGCCCTG 735
Qy 180 AspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMet 199
```

Db 1729 CGAGACCCCAAGGCCCAATGCCCCGTACACGGGGGGGGCTCAG 1776

Search completed: August 5, 2004, 21:21:04
Job time : 132 secs

```

736 CCGGGAGCCGA---GAGGCCCCGGCAATGTGGGTCTCTGGATCAGAGCTGGCCCTG 792
QY
200 GlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysValThrIlePhe 219
Db
793 CAGTGGGTGCAGGAGAACGTGGCAGCTTCCGGGGGTGACCCGACATCAGTGACGCTGTT 852
QY
220 GlyGluSerAlaGlySerMetSerValLeuLeuCyHisIleuIleTrpAsnAspGlyAspAsn 239
Db
853 GGGGAGAGCGCGGAGCGCTCGGTGGGCATGACCTGCTG----- 894
QY
240 ThrTrpLysGlyLysProLysPheArgAlaGlyIleMetGlnSerGlyAlaMetValPro 259
Db
895 TCCCGCCCGCCAGCCGGGCTGTTCCACAGGGCCGTGCTGCAGAGGGTGCC-----CCC 948
QY
260 SerAspProValAspGlyThrTrpGly-----AsnGluIle 271
Db
949 AATGGACCCCTGG---GCCACGGTGGGCATGGGAGAGGCCCGTGCAGGGGCCACGCGCTG 1005
QY
272 TyrAspLeuPheValSerSerAlaGlyCys-----GlySerAlaSer 285
Db
1006 GCCCACCTT-----GTGGGCTGTCTCCAGGGGCGACCTGGTGGGATGACACA 1053
QY
286 AspLysLeuAlaCysLeuArgSerAlaSerAspThrLeuLeuAspAlaThrAsnAsn 305
Db
1054 GAGCTGTGTAGCTGCTTCGGACACGACGACGAGTCTCTGTG-----AACCAC 1104
QY
306 ThrProGlyPheLeuAlaTy-SerSerLeu---ArgLeuSerTyLeuProArgProAsp 324
Db
1105 GAATGGCAGCTGTGCCTCAAGAAAGCTCTTCCGGTTCTCTCGTCTGTGTGTAGAT 1164
QY
325 GlyLysAsnIleThrAspAspMetTyLysLeuValArgAspGlyLysTyAlaSerVal 344
Db
1165 GGAGACTTCTCAGTGACACCCCGAGGCCCTCATCAACGGGGAGACTTCCACGGCCCTG 1224
QY
345 ProValIleIleGlyAspGlnAsnAspGluGlyThrIlePhe----- 358
Db
1225 CAGGTGCTGGTGGGTGGTGAAGGATGAGGCTCGTATTCTGTGTTTACGGGGCCCCA 1284
QY
359 GlyLeuSerSerLeuAsn-----ValThrAsnAlaGlnAlaArgAlaTyPheLys 376
Db
1285 GGCTTTCAGCAAGACAAACAGAGTCTCTCATCAGCGGGCGGAGTTCCTGGCCGGGGTGGG 1344
QY
377 GlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTy----- 394
Db
1345 GTGGGGGTCCCGAGTAAGTACCTGGCAGCGCGGCTGTGTGCTGCTGATTACACAGAC 1404
QY
395 -----ProGlnAsp-----IleThrGlnGlySerProPheAspThrGlyVal 408
Db
1405 TGGCTGCATCCCGAGGACCCCGCAGCCCTGAGGGAGGCCCTGAGCGATGTGTGGCGGAC 1464
QY
409 LeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPhe 428
Db
1465 CACAAATGCTGTGCCCC-----GTGGCCAGCTGGCTGGCGACTGGCTGCC 1512
QY
429 IleHisAlaArg-----ArgTyPheLeuAsnHisPheGlnGly----- 441
Db
1513 CAGGGTGCCCGGGTCTACGCCTACGCTTTGAAACACCGTGTCTCCAGGCTCTCTGGCCC 1572
QY
442 -----GlyThrLysTySerPheLeuSerLysGlnLeuSerGlyLeuProIle--- 457
Db
1573 CTGTGGATGGGGTGGCCCGCAGCGCTACGAGATCGAGTTCATCTTTGGGATCCCGCTGGAC 1632
QY
458 ---MetGlyThrPheHisAlaAsnAspIleValTrpGlnAspTyTrpLeuLeuGlySerGly 476
Db
1633 CCCTCTCGAAACTACACGCGCAGAGAGAAAATCTTCGCCAGCGACTGATG----- 1683
QY
477 SerValIleTyAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAla 496
Db
1684 -----CGATACTGGGCCAACTTTGCCCGCACAGGGGATCCCATGAGCCC 1728
QY
497 -----GlyLeuLeuValAsnTrpProLysTyThrSerSerSerGln 510

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 5, 2004, 20:33:42 ; Search time 515 Seconds
(without alignments)
5207.801 Million cell updates/sec

Title: US-09-943-857-4

Perfect score: 2864

Sequence: 1 SMNSRCPAGRLGVSPTAKLA.....DNFRTAGYDALMTNPSPFFV 547

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5	
Fgapop 6.0	Fgapext 7.0	
Delop 6.0	Delext 7.0	

Searched: 3222919 seqs, 2451570024 residues

Total number of hits satisfying chosen parameters: 6445838

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09943857/runat.29072004.090902.29857/app.query.fasta.1.711
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNIT8=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09943857 @CGN 1.1.354 @runat.29072004.090902.29857
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NRG SCORES=0 -WAIT -DSPBLOCk=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/2/pubpna/PTC_NEW_PUB.seq:
3: /cgn2_6/ptodata/2/pubpna/US05_NEW_PUB.seq:
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/2/pubpna/PTCUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
------------	-------	-------	--------	-------	-------------

1	2392.5	83.5	1532	10	US-09-943-857-3	Sequence 3, Appli
2	2034	71.0	1541	10	US-09-943-857-7	Sequence 7, Appli
3	2020	70.5	1548	10	US-09-943-857-5	Sequence 5, Appli
4	1903	66.4	1511	10	US-09-943-857-9	Sequence 9, Appli
5	1875.5	65.5	1469	10	US-09-943-857-1	Sequence 1, Appli
6	1156.5	40.4	1687	16	US-10-369-493-27923	Sequence 27923, A
7	456	15.9	2416	15	US-10-032-233-39	Sequence 39, Appli
8	456	15.9	2416	17	US-10-324-466-39	Sequence 39, Appli
9	456	15.9	2416	17	US-10-413-432-39	Sequence 39, Appli
10	455	15.9	2416	9	US-09-748-739A-7	Sequence 7, Appli
11	452	15.8	1725	9	US-09-810-861B-5	Sequence 5, Appli
12	452	15.8	2416	15	US-10-032-233-33	Sequence 33, Appli
13	452	15.8	2416	15	US-10-324-466-33	Sequence 33, Appli
14	452	15.8	2416	17	US-10-413-432-33	Sequence 33, Appli
15	452	15.8	5767	9	US-09-810-861B-3	Sequence 3, Appli
16	452	15.8	14446	9	US-09-810-861B-4	Sequence 4, Appli
17	451	15.7	1725	16	US-10-326-892-1	Sequence 1, Appli
18	451	15.7	2381	9	US-09-880-107-2271	Sequence 2271, Ap
19	451	15.7	2416	9	US-09-748-739A-5	Sequence 5, Appli
20	451	15.7	2416	9	US-09-748-739A-16	Sequence 16, Appli
21	451	15.7	2416	15	US-10-032-233-21	Sequence 21, Appli
22	451	15.7	2416	15	US-10-032-233-23	Sequence 23, Appli
23	451	15.7	2416	15	US-10-032-233-25	Sequence 25, Appli
24	451	15.7	2416	15	US-10-032-233-43	Sequence 43, Appli
25	451	15.7	2416	17	US-10-324-466-21	Sequence 21, Appli
26	451	15.7	2416	17	US-10-324-466-23	Sequence 23, Appli
27	451	15.7	2416	17	US-10-324-466-25	Sequence 25, Appli
28	451	15.7	2416	17	US-10-413-432-23	Sequence 23, Appli
29	451	15.7	2416	17	US-10-413-432-21	Sequence 21, Appli
30	451	15.7	2416	17	US-10-413-432-23	Sequence 23, Appli
31	451	15.7	2416	17	US-10-413-432-25	Sequence 25, Appli
32	451	15.7	2416	17	US-10-413-432-43	Sequence 43, Appli
33	451	15.7	2444	10	US-09-893-519A-112	Sequence 112, App
34	450	15.7	2416	9	US-09-748-739A-3	Sequence 3, Appli
35	450	15.7	2416	15	US-10-032-233-27	Sequence 27, Appli
36	450	15.7	2416	15	US-10-032-233-31	Sequence 31, Appli
37	450	15.7	2416	15	US-10-032-233-35	Sequence 35, Appli
38	450	15.7	2416	15	US-10-032-233-37	Sequence 37, Appli
39	450	15.7	2416	17	US-10-324-466-31	Sequence 31, Appli
40	450	15.7	2416	17	US-10-324-466-35	Sequence 35, Appli
41	450	15.7	2416	17	US-10-324-466-37	Sequence 37, Appli
42	450	15.7	2416	17	US-10-413-432-27	Sequence 27, Appli
43	450	15.7	2416	17	US-10-413-432-31	Sequence 31, Appli
44	450	15.7	2416	17	US-10-413-432-35	Sequence 35, Appli
45	450	15.7	2416	17	US-10-413-432-35	Sequence 35, Appli

ALIGNMENTS

RESULT 1
US-09-943-857-3
; Sequence 3, Application US/09943857
; Publication No. US20030124701A1
; GENERAL INFORMATION:
; APPLICANT: Shaw, Jei-Fu
; APPLICANT: Lee, Guan-Chiun
; APPLICANT: Tang, Shye-Jye
; TITLE OF INVENTION: RECOMBINANT CANDIDA RUGOSA LIPASES
; FILE REFERENCE: 08919-066001
; CURRENT APPLICATION NUMBER: US/09/943,857
; CURRENT FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1532
; TYPE: DNA
; ORGANISM: Candida rugosa
US-09-943-857-3

Alignment Scores: 7.83e-279 Length: 1532
Pred. No.: 2392.50 Matches: 502
Score:

Percent Similarity:	94.37%	Conservative:	1
Best Local Similarity:	94.18%	Mismatches:	9
Query Match:	83.54%	Indels:	26
DB:	10	Gaps:	16
US-09-943-857-4 (1-547) x US-09-943-857-3 (1-1532)			
QY	15	ProThrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAlaIleLeuAsnGlu	34
DB	2	CCACCGCAAGTCGGCAACGCGGACACCATCACCGGTCTCAACGCCATCATCAACGAG	61
QY	35	AlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArgPheLysAspPro	54
DB	62	GGTTCCTCGCATTCCTTTGCGGAGCGCGGTGGCAACCTCCGGCTTCAAGGACCT	121
QY	55	ValProFyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyProSerCysMet	74
DB	122	GTGCGGTACTCTGGCTCGCTCAACGGCCAGAAGTT-ACT---TACGGCCCG---TGATG	174
QY	75	GlnGlnAsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAlaLeuAspLeuVal	94
DB	175	CAGCAGAACCCCGAGGCGACGTTTGAAGAGAACCTTGGCAAGACGGCACTCGACTTGGT	234
QY	95	MetGlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsn	114
DB	235	ATGAGTCCCAAGGTGTTCCAGCGGTGCTTCCCGAGAGTGAAGTGCCTCACCATCAAC	294
QY	115	ValValArgProGlyThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePhe	134
DB	295	GTGTTGCGCGCGCGGCGCACCAAGCGCGGCGCCAACTCCCGGTCTGCTCGATCTTT	354
QY	135	GlyGlyGlyPheGluIleGlySerProThrIlePheProProAlaGlnMetValThrLys	154
DB	355	GGCGGTGGGTGTGAGATCGGACGCCCAACATCTTCCCTCCCGCCAGATGGTCACCAAG	414
QY	155	SerValLeuMetGlyLysHisIleIleHisValAlaValAsnTyrArgValAlaSerTrp	174
DB	415	AGTGTGCTCATGGGCAAGCC-ATCATCCACGTGGCGGCTCAACTACCGGTGTCCCTCGTG	473
QY	175	GlyPheLeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAsp	194
DB	474	GGGTTCCTGGTGTGATGATCATCAAGCGCGGAGCGCGGGAACGCGGCTTGAAGGAC	533
QY	195	GlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSer	214
DB	534	CAGCGTTTGGCATGAGTGGTGGCAGACACATTTGCCGGGTTCGGCGGCGACCCGAGC	593
QY	215	LysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCysHisIleLeuTrp	234
DB	594	AAGGTGACATC-TTTGGCGAG---GCGGCGACATGTCCGTGTTGTGCCACTCATCTGG	649
QY	235	AsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSer	254
DB	650	AACGACGGCGCAACACGTCACAAAGGCGAAGCGGTGTTCCGCGCGGCGCATGTGAG--	706
QY	255	GlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGluIleTyrAspLeu	274
DB	707	GGAGCCATGTGCCG---GACCCGGTGGACGCGGACGTACGGCAACGAGATCTACGACCTC	763
QY	275	PheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCysLeuArgSerAla	294
DB	764	TTTGTCTCGAGTGTGCTGTGCGAGCGCGCAGCGCAAGCTCGCGTGTGGCGAGTGGC	823
QY	295	SerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeuAlaTyrSerSer	314
DB	824	---AGCGACACCTTGTCTGATGCGCACCAACACACTCTCGGGTCTTGGCGTACTCTCG	880
QY	315	LeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThrAspAspMetTyrLys	334
DB	881	TTGCGGTG-----TACTCCCGCGCGCGGCAAGACATCACCGATGATGTACAAG	934
QY	335	LeuValArgAspGlyLysTyrAlaSerValProValIleLeuGlyAspGlnAsnAspGlu	354

```

Db      7  GCGACGGTCGCGCAACGGCGCACCATCACCGGTCTCAACGGCATCATCAACGAGCGGTC 66
Qy      37  LeuGlyIleProPheAlaGluProValGlyAsnLeuArgPheLysAspProValPro 56
Db      67  CTGGCATTCCTTTGCGGAGCGCGGTGGCAACCTCCGCTTCAGAGACCTTGTGGCG 126
Qy      57  TyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyProSerCysMetGlnGln 76
Db      127  TACCGTGGGTCTCTCAACGGGTCAATCTTCAACGGGTACGGTCCG---TGATCCAGCAG 183
Qy      77  AsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGln 96
Db      184  AACCCGAGGACCTACGAGGAGAACCTCCCAAGGTGGCGCTTGACTTGGTATGACG 243
Qy      97  SerLysValPheGlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsnValVal 116
Db      244  TCCAAAGGTGTTCAGGCTGTCTCCCCCAACGAGGAGACTGCTCACCATCAAGTGGTG 303
Qy      117  ArgProGlyThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyGly 136
Db      304  CGGCGCGCGGCGCAAGCGGCGCGCAACCTCCCGGTCTGCTCTCGATCTTTGGCGGT 363
Qy      137  GlyPheGluIleGlySerProThrIlePheProGlnMetValThrLysSerVal 156
Db      364  GGGTTTGAGATCGGACGCGCCACCATCTTCCCTCCCGCTCAGATGGTCTCCAAGAGTGTG 423
Qy      157  LeuMetGlyLysHisIleLeHisValAlaValAsnTyrArgValAlaLaserTrpGlyPhe 176
Db      424  CTCATGGGCGGCC-ATCATCCAGTGGCGTCAACTACCGCTGGCGTGGCTCTTTGGTTTC 482
Qy      177  LeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArg 196
Db      483  TTGGCGCGTCCGACATCAAGCGCGGCGGAGCTCCAAATCGCGGCTCAAGGACCGCGC 542
Qy      197  LeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysVal 216
Db      543  TTGGGATGCGTGGTGGGAGACACATTTGCCGGGTTCGGCGGCGACCGCGACCAAGGTG 602
Qy      217  ThrIlePheGlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAsp 236
Db      603  ACCATCTTTGGAG---GGGCGCAGCATGTCGCTGTGTGGCCACCTTCTCGAATGGC 658
Qy      237  GlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAla 256
Db      659  GCGCAACACGATACAGGCGCAAGCGGTGTTCCGCGCGGCGCATCATGCG---GGAGCC 715
Qy      257  MetValProSerAspProValAspGlyThrTyrGlyAsnGluIleTyrAspLeuPheVal 276
Db      716  ATGGTGCCG---GACCCGGTGGACGCGACCTATGGACC-CAATCTATGACACGTTGGTG 771
Qy      277  SerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCysLeuArgSerAlaSerSer 296
Db      772  GCTACG---GGCTGCAGCAGTGCACACACAGCTTGGCTGTGGTGGTCTT---ACT 825
Qy      297  AspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeuAlaTyrSerSerLeuArg 316
Db      826  CAGGCATTCTCGATGCCACCAACGACGCGCTGGGTCTTGTGCTACACCTCGTTGGCG 885
Qy      317  LeuSerTyrLeuProArgProAspGlyLysAsnIleThrAspAspMetTyrLysLeuVal 336
Db      886  TTG-----TACTCCGGCGCGCGCGCAACATCACCGATGACATGTACAAAGTTGGTA 939
Qy      337  ArgAspGlyLysTyrAlaSerValProValIleIleGlyAspGlnAsnAspGluGlyThr 356
Db      940  CGCGAGCGCAAGTATGCAAGCGTTCCGCTGATCATTTGGCGACACAGAACGAGGCGCTTC 999
Qy      357  IlePheGlyLeuSerSerLeuAsnValThrThrAsnAlaGlnAlaArgAlaTyrPheLys 376
Db      1000  TTGTTTGT-CTC-----TTGAACACCAACCGAGCGCGCGAGGCGATACCTCAGA 1052
Qy      377  GlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyrProGln 396

```

```

Db      1053  AAG---TTTCATCCAGCGCACCGCGGATATACCGCATTTGAAGCGCGGTACCCAGC 1109
Qy      397  AspIleThrGlnGlySerProPheAspThrGlyValLeuAsnAlaLeuThrProGlnPhe 416
Db      1110  GATGTACCCAGGGT---CGGTTCCGACACGGGCAATCTCAACGCGCTTACACCCAGCTC 1166
Qy      417  LysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHisAlaArgAtgTyrPheLeu 436
Db      1167  AAGCGGATCAATGCTGCTTGGGACCTCACCTTTACCTCTCGCGCGCTACTTCTCTC 1226
Qy      437  AsnHisPheGlnGlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeuPro 456
Db      1227  AACCACTACACCGGTGGTCCCAAGTACTCGTTCTCTC---AAGCAGCTT---GGTTG-CCC 1279
Qy      457  IleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAspTyrLeuLeuGlySerGly 476
Db      1280  ATTCTCGGAGC-TTCCACGCGGAACGACATTTGTGTGGCAGCACTTTTGTGGCGCGGC 1338
Qy      477  SerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAla 496
Db      1339  AGCGTCATCTACAAACAGCGTTTCATCGCTTTGCCACCGACTTGGACCCCAACACCGCG 1398
Qy      497  GlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGlnSerGlyAsnAsnLeuMet 516
Db      1399  GGC---TTGGTGCAGTGGCGCAAGTC-ACC---AGCAGCCAGCGGCGGACAACTTGTG 1451
Qy      517  MetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGlyTyrAsp 536
Db      1452  CAGTCAGTCCCTTGGGCTTGTACACGCGCAAGGACAACTTCCGCGACCGCGGCTACAAC 1511
Qy      537  AlaLeuMetThrAsnProSerSerPhePheVal 547
Db      1512  GCTTTGTTTGGCGACCGG---CACTTTTTCGTG 1541

```

RESULT 3

```

US-09-943-857-5
; Sequence 5, Application US/09943857
; Publication No. US20030124701A1
; GENERAL INFORMATION:
; APPLICANT: Shaw, Jei-Fu
; APPLICANT: Lee, Guan-Chiun
; APPLICANT: Tang, Shye-Jye
; TITLE OF INVENTION: RECOMBINANT CANDIDA RUGOSA LIPASES
; FILE REFERENCE: 08919-066001
; CURRENT APPLICATION NUMBER: US/09/943,857
; CURRENT FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1548
; TYPE: DNA
; ORGANISM: Candida rugosa
US-09-943-857-5

```

Alignment Scores:

```

Pred. No.:      1e-233      Length:      1548
Score:          2020.00      Matches:      422
Percent Similarity: 87.12%      Conservative: 38
Best Local Similarity: 79.92%      Mismatches: 57
Query Match:      70.53%      Indels:      17
DB:              10         Gaps:      9

```

US-09-943-857-4 (1-547) x US-09-943-857-5 (1-1548)

```

Qy      15  ProThrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAlaIleIleAsnGlu 34
Db      2  CCACATGCGCAGCTCGCCAAACGGCGACACCATCACCGGTCTCAACGCCATCATCAACGAG 61
Qy      35  AlaPheLeuGlyIleProPheAlaGluProValGlyAsnLeuArgPheLysAspPro 54
Db      62  GCGTTCTCCGTATTTCCCTTTGCTCAGCGCGCGTGGGCAACCTCCGCTTCAAGCGGCT 121
Qy      55  ValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyProSerCysMet 74

```

122 GTCCGCTACGCGCTCTCAATGGTCAGAGTTTACTGTATGG---CCCTG-TGCATG 177
 75 GlnGlnAsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAlaLeuAspLeuVal 94
 178 CAGATGAACCAATTGGGCAACTGGGACTCCTCGCTTCCCAAGGCTGCATCAACTTG--- 234
 95 MetGlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsn 114
 235 ATGCAGTCCCAAGCTCTTCCAGCGCGTCTTCCCTAACCGCGAGGACTGTCTCACCATCAAC 294
 115 ValValArgProProGlyThrLysAlaGlyAlaValAsnLeuProValMetLeuTriPhe 134
 295 GTGGTGGG---CCGGGACCAACCGCGGTGTCACCTCCCGGTGATGTTGTTG 351
 135 GlyGlyGlyPheGluIleGlySerProThrIlePheProAlaGlnMetValThrLys 154
 352 GCGCGCGGTTTGAGGTTGGCGGCTCCAGTCTCTTCCCTCCGACACAGATGATCACCGCC 411
 155 SerValLeuMetGlyLysHisIleIleHisValAlaValAsnTyrArgValAlaSerTrp 174
 412 AGCGTCTTATGGGCAACGCCATCATCCAGTGAAGTGAACCTACCGGTGCTTCGTGG 471
 175 GlyPheLeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAsp 194
 472 GGGTCTTGGTGTCTCCAGACATCAACCGCGGAGGCGGAGCGCGGTTTGACGAC 531
 195 GlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSer 214
 532 CAACGCTTGGTGTTCAGTGGTGGCGACACATTCGCGGGTTCGCGCGGACCGCTCC 591
 215 LysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrp 234
 592 AAGGTGACCATCTTTGGTGAGG--GCGGGCAGCATGTCCGTAATGTGTCAGCTCTCTGG 649
 235 AsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSer 254
 650 AACGACGCGGCAACACAGTACACCGCAACGCTGTTCCGTCGCGCCCATCATGACG-- 706
 255 GlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGluIleTyrAspLeu 274
 707 GGGGCGCATGTTGCCCG--GACCGGTGGATGGCGCTACGCGCAGCAGATCTACGACCC 764
 275 PheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCysLeuArgSerAla 294
 765 GTGGTGTCTACGCGCGCTGTGCAGTGCACGCAAGCTCGCTGCTTGGCA--GCA 822
 295 SerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeuAlaTyrSerSer 314
 823 TGAACGACAAACTCTTCCAGGCGCACCGACACTCGGGGGCTTGGCGTACCCCTCG 882
 315 LeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThrAspAspMetTyrLys 334
 883 TTGGCGTGG--TTTCTCCGCGCGCGCGGACCTTCATCAGCATGACATGTTCAG 940
 335 LeuValArgAspGlyLysTyrAlaSerValProValIleGlyAspGlnAsnAspGlu 354
 941 TTGGTTCGCGACGCGCAAGTGTCCACGTTCCGCTGATCATTTGGCGACCCAGACGACGAG 1000
 355 GlyThrIlePheGlyLysSerSerLeuAsnValThrThrAsnAlaGlnAlaArgAlaTyr 374
 1001 GGCACAGTGTTCGCTG-----TTGAACGTGACTACCGATGCTCAGGCGACCGCATG 1054
 375 PheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaIaTyr 394
 1055 TTCAAGGAA---TTCATCCACGCGCAGCAGCGGAGATGCACACTTGTATGGCGGCTAC 1111
 395 ProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsnAlaLeuThrPro 414
 1112 CCAGCGACATCAACCGAGGT---CCGTTCCACACCGGCGATCTTCACGCGCATCACCCCG 1168
 415 GlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHisAlaArgArgTyr 434

1169 CAGTTCAAACCGGATTGTCAGCGGTGCTGGTGACCTTCGGTTCACTCTCCCGCGGCTAC 1228
 435 PheLeuAsnHisPheGlnGlyGlyThrLysTyrSerPheLeuSerLysGlnLeuSerGly 454
 1229 TTCTCTACACCATTCCTCAGGCGGCGCACCAAGTACTCGTTCC--TCGAAGCAGCTT---GGG 1283
 455 LeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAspTyrLeuLeuGly 474
 1284 TTGCGGTGATTTGGCACCCACCGCACGACATTGTGTGGCAGGACTTTTGGTGAGC 1343
 475 SerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsn 494
 1344 CACAGCAGCGCGGTGTACACACCGCTTATTCCTTTGCCACACGCTCGACCCGAC 1403
 495 ThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGlnSerGlyAsnAsn 514
 1404 AAGCGCGTGTCTTGTGTAAGTGGCCCAAGTACACGACGACG-----CAGGGCAACAC 1457
 515 LeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGly 534
 1458 TTGTTGCAGATCAACGCGCTTGGGCTTGTACACCGGCAAGGACAACTTCCCGCACCGCTGGC 1517
 535 TyrAspAlaLeuMetThrAsnPro 542
 1518 TACGACGCTGTGTACCAACCCG 1541
 RESULT 4
 US-09-943-857-9
 ; Sequence 9, Application US/09943857
 ; Publication No. US20030124701A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Shaw, Jai-Fu
 ; APPLICANT: Lee, Guan-Chiun
 ; APPLICANT: Tang, Shye-Jye
 ; TITLE OF INVENTION: RECOMBINANT CANDIDA RUGOSA LIPASES
 ; FILE REFERENCE: 08919-066001
 ; CURRENT APPLICATION NUMBER: US/09/943,857
 ; CURRENT FILING DATE: 2001-08-31
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 9
 ; LENGTH: 1511
 ; TYPE: DNA
 ; ORGANISM: Candida rugosa
 US-09-943-857-9
 Alignment Scores:
 Pred. No.: 1,41e-219 Length: 1511
 Score: 1903.00 Matches: 433
 Percent Similarity: 86.84% Conservative: 29
 Best Local Similarity: 81.39% Mismatches: 43
 Query Match: 66.45% Indels: 38
 DB: 10 Gaps: 20
 US-09-943-857-4 (1-547) x US-09-943-857-9 (1-1511)
 QY 17 AlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAlaIleIleAsnGluAlaPhe 36
 DB 7 GCCACGCTCGCCCAACGGCGACACCATCACCGGTCTCAACGCCATCATCAACGAGCGGCTC 66
 QY 37 LeuGlyIleProPheAlaGluProProValGlyAsnLeuArgPheLysAspProValPro 56
 DB 67 CTCGGCATTCCTTTCCGAGCGCGGTGGGCACTCCCGCTTCAAGACCCCGTGGCG 126
 QY 57 TyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyProSerCysMetGlnGln 76
 DB 127 TACTCCGCTCGCTCGATGCCAGAGTTTCAC-----TACGGCCG---TGCATGACGAC 179
 QY 77 AsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGln 96
 DB 180 AACCCCGAGGCGACCTTACGAGGAGAACCTCCCGGAGGCGCTCGCTCGATGGTGGATGAC 239
 QY 97 SerLysValPheGlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsnValVal 116


```

Db      240  TCCAGAGTGTGTGAGGCGGTG-----CCGAGCGAGGACTGTCTCACCATTCAACGCTGGTG 293
Qy      117  ArgProGlyThrLysAlaGlyAlaAsnLeuProValMetLeuTrrPheGlyGly 136
Db      294  CGCGCGCGCGGACCAAGCGGTGCCAACCTCCGGGTGATGCTCTGATCTTTGGCGGC 353
Qy      137  GlyPheGluLeuGlySerProThrIlePheProProAlaGlnMetValThrLysSerVal 156
Db      354  GGGTTTGGAGTGGGTGGCAGCACCCTTCCTCCCGCCGAGATGATCACCAGAGCAT 413
Qy      157  LeuMetGlyLysHisIleIleHisValAlaValAsnTyrArgValAlaSerTrpGlyPhe 176
Db      414  GCCATGGCAAGGCCATATCCAGTGGAGCGTCAACTACCGCGTGTGCTGTGGGGTTC 473
Qy      177  LeuAlaGlyAspAlaGlyLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArg 196
Db      474  TTGGCTGGCGACGAGATCAAGCGCGAGGCGAGTGCACACGCCGGTTTGAAGGACCAAGCG 533
Qy      197  LeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysVal 216
Db      534  TG-GGCATGAGTGGGTGGCGGACCAACATTCGCGCGTTCGGCGCGACCCGACCAAGGTG 592
Qy      217  ThrIlePheGlyGluSerAlaGlySerMetSerValLeuCysHisIleLeuTrpAsnAsp 236
Db      593  ACCATCTTTGGCGAG---GGCGGAGCATGTGCGTCAATGTGCCACATTTCTCGAAGCAC 649
Qy      237  GlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAla 256
Db      650  GCGCAACACAGTACAGGCGACAGCGCTCTTCGGCGCGGCAATCATGCAG---GGGGCC 706
Qy      257  MetValProSerAspProValAspGlyThrTyrGlyAsnGluLeuTyrAspLeuPheVal 276
Db      707  ATGGTC---CGGACGCG-GTGGACGGCTC-TACGGCAACGAGATCTTTGACCTTTGGCG 761
Qy      277  SerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCysLeuArgSerAlaSerSer 296
Db      762  TCGAC-GCGGCTGGCGGCGAGCGCGACGACCAAGCTTCGCGTGTGGCGGTGTG---AGC 817
Qy      297  AspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeuAlaTyrSerSerLeuArg 316
Db      818  GACACCTTGGAGGACCGCCACCAACACCTTGGGTCTTGGCGTACTCTCTGTTGGCG 877
Qy      317  Leu-SerTyrLeuProArgProAspGlyLysAsnIleThrAspAspMetTyrLysLeuVa 336
Db      878  TTGTACT-----CCCGCGCGCGCGCGCGTGAACATCACCGACGACATGTGC---CTTG 928
Qy      336  lArgAspGlyLysTyrAlaSerValProValIleIleGlyAspGlnAsnAspGluGlyTh 356
Db      929  TCGGAGGGCAAGTATGCAT-----CCTGTGATCATCGCGGACCCAGAACGACGAGGCGAC 984
Qy      356  rIlePheGlyLeuSerSerLeuAsnValThrThrAsnAlaGlnAlaArgAlaTyrPhely 376
Db      985  CTTCTTTGCG-----ACCTTGACCTGACCGGAT-----GCCAGCGCCGAGTACT 1032
Qy      376  sGlnSerPheIleHisAlaSerAspAlaGluLeuAspThrLeuMetAlaAlaTyrProgl 396
Db      1033  TCAGCAGTTTGTCCAGCGCGAGCGCGAGTGC-GACACGTTGATGACGCGGTACCCC--- 1089
Qy      396  nAspIleThrGlnGlySerProPheAspThrGlyValLeuAsnAlaLeuThrProGlnPh 416
Db      1090  -GACATCACCCA-GGG---CCGTGTGACACGGGTTT-CTCAACGCGCCCTCACCGCGAGTT 1143
Qy      416  eLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHisAlaArgTyrPheLe 436
Db      1144  CAAGAGATC---CGGTGCTCTGGCGACCTTGTCTACGC-----CGTGGCTACTTCT 1193
Qy      436  uAsnHisPheGlnGlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeuPr 456
Db      1194  CAACCACTACACCGCGCGGACCAAGTACTCATTCCTC---AAGCAGCTC---GGCTTGC 1247
Qy      456  oIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAspTyrLeuLeuGlySerGl 476

```

```

Db      1248  GGTGCTCGGAAGCTTCCACTCCACGACATTGTCTTCCAGGACTACTTGTGGGAGCGG 1307
Qy      476  ySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAl 496
Db      1308  CTGCTCATCTACAAACACGCTTTCATTGGCTTTGCCACGAGCTTGGACCCCAACCGC 1367
Qy      496  aGlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGlnSerGlyAsnAsnLeuMe 516
Db      1368  GGGGTGTTGGTGAAGTGGCCGAGTACACAGCAGC-----CAGGGCAACAACCTTGAT 1421
Qy      516  tMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGlyTyrAs 536
Db      1422  GATGATCAACGCTTGGCTTTGATACACCGCAAGGACAACTC-CGCACCGCGCGCTACGA 1480
Qy      536  pAlaLeuMetThrAsnProSerSerPheVal 547
Db      1481  CGCGTTGTTCTTCCAAACCGC---CCGTTCTTTGTG 1511

RESULT 5
US-09-943-857-1
; Sequence 1, Application US/09943857
; Publication NO. US20030124701A1
; GENERAL INFORMATION:
; APPLICANT: Shaw, Jai-Fu
; APPLICANT: Lee, Guan-Chiun
; APPLICANT: Tang, Shye-Jye
; TITLE OF INVENTION: RECOMBINANT CANDIDA RUGOSA LIPASES
; FILE REFERENCE: 08919-066001
; CURRENT APPLICATION NUMBER: US/09/943,857
; CURRENT FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1469
; TYPE: DNA
; ORGANISM: Candida rugosa
US-09-943-857-1

Alignment Scores:
Pred. No.: 2,87e-216 Length: 1469
Score: 1875.50 Matches: 394
Percent Similarity: 80.80% Conservative: 31
Best Local Similarity: 74.90% Mismatches: 61
Query Match: 65.49% Indels: 44
DB: 10 Gaps: 16

US-09-943-857-4 (1-547) x US-09-943-857-1 (1-1469)
Qy      19  LeuAlaAsnGlyAspThrIleThrGlyLeuAsnAlaIleIleAsnGluAlaPheLeuGly 38
Db      12  CTCGCCAACGGCGACACCATCACCGTCTCAACGCCATTGTCAACGAAAGTTTCTCGGC 71
Qy      39  IleProPheAlaGluProProValGlyAsnLeuArgPheLysAspProValProTyrSer 58
Db      72  ATACCGTTTGGCGAGCGCGCGTGGCA--CTCCGCTTCAAG-----111
Qy      59  GlySerLeuAsnGlyGlnLysPheThrSerTyrGlyProSerCysMetGlnGlnAsnPro 78
Db      112  -----CTCAACGGCGACGAGTTTACC---TACGCGCCG---TGCATGCGAGTGAACCT 159
Qy      79  GluGlyThrPheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLys 98
Db      160  ATGGGCTCGTTTCA-----TTGGTGCTCCAGTCCAG 191
Qy      99  ValPheGlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgPro 118
Db      192  ATCTTCCAAGTGTGCTTCCCAACGACGAGGACTGTCTCACC-----233
Qy      119  ProGlyThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyGlyPhe 138
Db      234  -----ACCAGGCGAGTGTGCTCCCGGTGATGTCTCGATCTTTGGCGTGGT 287
Qy      139  GluIleGlySerProThrIlePheProAlaGlnMetValThrLysSerValLeuMet 158

```

288 GAGCTTGGCGGCTCCAGCCTTTTCCAGGAGACAGATGTTGGCCAGAGCGTGTCTATG 347
 159 GlyLysHisIleIleHisValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAla 178
 348 GGTAAACCGGTGATCCACGTCGAGATGAATACCGCGTGGCGTTCATGGGGTCTTGGCC 407
 179 GlyAspAspIleIleHisValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAla 198
 408 GGCCCCGACATCCAGAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 467
 199 MetClnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysValThrIle 218
 468 ATGACGATGGGTGGGGGAGCAACATCTGCTGGTGGGGGAGGAGGAGGAGGAGGAGGAG 527
 219 PheGlyGluSerAlaGlySerMetSerValLeuGlyHisIleValAlaSerTrpGlyPhe 238
 528 TACGGCGAG--GGGGGAGCATGTCAGCGTTTGTGCACCTTGTGTGGAACGAGCGGAC 584
 239 AsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetVal 258
 585 AACACGTCACACGCGCAAGCGGTGTCGCGCGCCCATCATGCGAG--GGCTGATGGTG 641
 259 ProSerAspProValAspGlyThrTyrGlyAsnGluIleTyrAspLeuPheValSerSer 278
 642 CCG--GACCCGGTGGAGCGGACGTCGCGCAGCCGAGATCTACACCGAGTGTGGCGTCT 698
 279 AlaGlyCysGlySerAlaSerAspLysLeuAlaCysLeuArgSerAlaSerSerAspThr 298
 699 GCGGGTGTGGCAGTGGCGAGCAGCAAGCTCGCGTGTGCGCGGCTTTCTCAGGACAG 758
 299 LeuLeuAspAlaThrAsnAsnThrProGlyPheLeuAlaTyrSerSerLeuArgLeuSer 318
 759 TTGTACACGAGCGACGAGCAGCGCGCGGTGTGGCGTACCGCGTGTGGCGGTG-- 815
 319 TyrLeuProArgProAspGlyLysAsnIleThrAspAspMetTyrLysLeuValArgAsp 338
 816 TATCTCCCGCGCGCGCGGACCTTCATCCCGCAGCATGTATGCTTGGTGGCGGAC 875
 339 GlyLysTyrAlaSerValProValIleGlyAspGlnAsnAspGluGlyThrIlePhe 358
 876 GGCAAGTACGACACGTCGCGGTGATCATCGCGCAGCAGAACGAGCGGCGCATTTGTT 935
 359 GlyLeuSerSerLeuAsnValThrThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSer 378
 936 GGGCTC-----TTGAACGTGACACAGATGCTCAGGCGACGCGGCGTACTTCAAGCAG-- 986
 379 PheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyrProGlnAspIle 398
 987 TTCATCCACGCGCGGATGCGGAGATCGACAGTTGTATGGCGCGGTACACCGAGCATC 1046
 399 ThrGlnGlySerProPheAspThrGlyValLeuAsnAlaLeuThrProGlnPheLysArg 418
 1047 ACCCAGGGT--CCGTTCGACACCGCGCATCTTCAATGGCCATCACCCTCGCGTCAACCG 1103
 419 IleSerAlaValLeuGlyAspLeuAlaPheIleHisAlaArgArgTyrPheLeuAsnHis 438
 1104 ATC--GGGTTCCTTGGCGACCTTGGTTCACGCTTGGCGTGGCGTACTTCTCAACTAC 1160
 439 PheGlnGlyGlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeuProIleMet 458
 1161 TACCAGGGCGGCGACCAAGTACTCGTCTC---AAGCAGCTT--GGGTGGCGCGTCTTG 1213
 459 GlyThrPheHisAlaAsnAspIleValTrpGlnAspTyrIleLeuGlySerGlySerVal 478
 1214 GGCACCTTCCAGCGCAACGACATCATCTGCGAGGACTACTTGGTGGCGAGCGGCGGTG 1273
 479 IleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAlaGlyLeu 498
 1274 ATCTACAAACACGCGTTCATTTGGTTTGGCAGACCTCGACCGCGACGAGGCGGCTTG 1333
 499 LeuValAsnTrpProLysTyrThrSerSerGlnSerGlyAsnAsnLeuMetIle 518

1334 TGGACCAACTGGCCCACTGATACACGAGCAG-----GGCAACAACCTTGATGCAGATC 1386
 519 AsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGlyTyrAspAlaLeu 538
 1387 AACGGCTTGGGTGTGTACACCGGAGGACAACTTCGCGCGGATGGTACAGGCGCTC 1446
 539 MetThrAsnProSerSer 544
 1447 TTTTCCAAACCGCTTCT 1464
 RESULT 6
 US-10-369-493-27923
 ; Sequence 27923, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; PRIOR FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 27923
 ; LENGTH: 1687
 ; TYPE: DNA
 ; ORGANISM: Neurospora crassa
 US-10-369-493-27923
 Alignment Scores:
 Pred. No.: 4,02e-129 Length: 1687
 Score: 1156.50 Matches: 247
 Percent Similarity: 60.39% Conservative: 96
 Best Local Similarity: 43.49% Mismatches: 175
 Query Match: 40.38% Indels: 50
 DB: 16 Gaps: 10
 US-09-943-857-4 (1-547) x US-10-369-493-27923 (1-1687)
 QY 13 SerValProThrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAlaIleIle 32
 DB 1 GCGGCGCGGACTGTCCACCATTCACCGGT---ACGATTGGGTGGTAAACGGCATCCTC 57
 QY 33 AsnGluAlaPheLeuGlyIleProPheAlaGluProValGlyAsnLeuArgPheLys 52
 DB 58 ACTGAGGCGCTTCAACGGAATCCCTTACGCGCTTCTCCGACCGGCAACCTTCGCTCAAG 117
 QY 53 AspProValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyProSer 72
 DB 118 CTTCCCGTGAGACTTAAGTCTCTCTGGGTCTTTGGTCTTGTATGCTCTGGCATCGCCCTGCT 177
 QY 73 CysMetGlnGlnAsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAlaLeuAsp 92
 DB 178 TGCCCCCAGTCTCTTGTGACACCTCGTGAACGAGTTTCTGCTCAG---GTTATCGAT 234
 QY 93 LeuValMetGlnSerLysValPheGlnAlaValLeuProGlnSerGlyAspCysLeuThr 112
 DB 235 AAGATCGTTAAACGACGAGCTTTTCAAGACTATCTACCTCAAGCTCAAGGAGGACTGCTTGACC 294
 QY 113 IleAsnValValArgProGlyThrLysAlaGlyAlaAsnLeuProValMetLeuTrp 132
 DB 295 ATCTCGGTCACTGCTGCTCCCAAGGCGACCAAGGCTGGTGAAGTCCCGCTCTCTTCTGG 354
 QY 133 IlePheGlyGlyPheGluIle----- 140
 DB 355 ATCTTTGGTGGTGGTTTCCGAAGTGAGAAATCCAGCTTATATACGCGATGTAATGAACAAG 414
 QY 141 -----GlySerProThrIlePheProAlaGlnMetValThrLys 154

Db 415 TGCTAAACTTCACAGCTCGGATCGCGTCCATGTACGATGCGCTCCCTAGTACCACAA 474
QY sSerValLeuMetGlyLysHisIleHisValAlaValAsnTyrArgValAlaSerTr 174
Db 475 CGCTATCAACATGGTAAAGCGTACGTCTACGTTCCCGTCACTACCGTGTGCGGTGTT 534
QY pGlyPheLeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAs 194
Db 535 TGGTTTCATGCCGGAAGAGATCTTAAGACCGCTCTTCCAACTTTGGGTACCTTGA 594
QY pGlnArgLeuGlyMetGlnTTPValAlaAspAsnIleAlaGlyPheGlyAspProSe 214
Db 595 CCAGCGATGGCGTCCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 654
QY rLysValThrIlePheGlyGlySerAlaGlySerMetSerValLeuCysHisLeuLeTr 234
Db 655 CAAGGTCACTATCTGGGCGAGTCCGCGGTGGTCCATGTCCGTTTCAACACAGATGTCTCT 714
QY pAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSe 254
Db 715 CTATGACGGTGACACACGTACACACGCGCAAGCCCTTTTCCGTGGCGCATCATGAATC 774
QY rGlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGluLeuTyrAspLe 274
Db 775 TGGTTTCATGCTCCCGCGCGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 834
QY uPheValSerSeraGlyCysGlySerAlaSerAspLysLeuAlaCysLeuArgSerAl 294
Db 835 CGTGTCAAGAACCGCGGTCTCTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 894
QY aSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeuAlaTyrSerSe 314
Db 895 TCCTACGAGACTTTTCTCAAGCGCGTAACTCCGTGCTGGGATTTCTGTGTACAATC 954
QY rLeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThrAspAspMetTyrLy 334
Db 955 CGTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1014
QY sLeuValArgAspGlyLysTyrAlaSerValProValIleIleGlyAspGlnAsnAspGl 354
Db 1015 GCTCATGCTCGTAAAGTACCGCGCGTCCCATGTATCATCGCGCATCAAGAGGATGA 1074
QY uGlyThrIlePheGlyLeuSerSerLeuAsnValThrAsnAlaGlnAlaArgAlaTy 374
Db 1075 GGGCACTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1134
QY rPheLysGlnSerPheIleHis--AlaSerAspAlaGluIleAspThrLeuMetAlaAl 393
Db 1135 CCTCAAGGATATCTTCTCAACGACGCCGAGTGGCAGATTAAGTCTCTCTCTCTGAC 1194
QY aTyrProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsnAlaLeuTh 413
Db 1195 CTACAGTACCTTATCTCGCGCGTCTCGCTTTGGCAGCGGCTCTTCAACAGATTTA 1254
QY rProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHisAlaArgAr 433
Db 1255 CCCCAGCTTCAAGCGCTGGCGCCATCTTGTGGATCTCTATCTTCAACCTCAGCGCGCG 1314
QY gTyrPheLeuAsnHisPheGlnGlyThr-----LysTyrSe 446
Db 1315 CATCTTCTCGAC-----GCCGCCACCACTCTCAACCCCTCGGTGCGCGCTGGTC 1365
QY rPheLeuSerLysGlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIl 466
Db 1366 GTATCTTGGTCTTACACTTTGGACACCCATCTTGGACCTTTCAGCGCTCCCGATAT 1425
QY e-----ValTrpGlnAspTyrLeuLeuGlySerGlySerValIleTyrAsnAsnAlaPh 484
Db 1426 CTGCAAGGTGTTCTACGGATCTCTGCCCACTACGCCGACCAAGCATCCAGTCTTACTA 1485
QY eIleAlaPheAlaThrAspLeuAspProAsnThrAlaGly----- 497

Db 1486 GCCCAACTTTGTTTACAACTTGACCCCAACGACGCTCCGCTGGCAGCTTCTCTAAGAG 1545
QY 498 ----LeuLeuValAsnTTPProLysTyrThrSerSerGlnSerGlyAsnAsnLeuMe 516
Db 1546 CAAGGTCAAGCGGATTTGGCGCAATGGCAGAGGAGAGAAG----- 1588
QY 516 tMetIleAsnAlaLeuGlyLeuTyrThrGly-----LysAspAsnPheArgThrAlaGl 534
Db 1589 -CTGGTCCAGTCTCTTTTCGGACTATGCGGATATCTTCCGGATGATTTCCGCTCTGATTC 1647
QY 534 yTyrAspAlaLeuMetThrAsn 541
Db 1648 GTATAACTGGATTAAAGGCTAAT 1669
RESULT 7
US-10-032-233-39
; Sequence 39, Application US/10032233
; Publication No. US20030153062A1
; GENERAL INFORMATION:
; APPLICANT: Watkins, Jeffrey D.
; APPLICANT: Hancock, James D.
; TITLE OF INVENTION: Butyrylcholinesterase Variants with
; FILE REFERENCE: P-IX 4642
; CURRENT APPLICATION NUMBER: US/10/032,233
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 2416
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Butyrylcholinesterase variant
; NAME/KEY: CDS
; LOCATION: (214)...(1935)
US-10-032-233-39
Alignment Scores:
Pred. No.: 4,62e-44 Length: 2416
Score: 456.00 Matches: 160
Percent Similarity: 45.17% Conservative: 83
Best Local Similarity: 29.74% Mismatches: 185
Query Match: 15.92% Indels: 110
DB: 15 Gaps: 26
US-09-943-857-4 (1-547) x US-10-032-233-39 (1-2416)
QY 22 GlyAspThrIleThrGlyLeuAsnAlaIleIleAsnGluAlaPheLeuGlyIleProPhe 41
Db 277 GGTGGCAGGTAACA-----GCCCTTCTTGAATTCCTAT 312
QY 42 AlaGluProValGlyAsnLeuArgPheLysAspProValProTyrSerGlySerLeu 61
Db 313 GCACAGCCACCTCTGGTAGACTTCGATTCAAAAGCCAGCTCTGACCAAGTGTCT 372
QY 62 AsnGlyGlnLysPheThrSerTyrGlyProSerCysMetGlnGlnAsnProGlyThr 81
Db 373 GATATTGGATGCCCAAAATATGCAAAATCTTGTCTGT---CAGAACATAGATCAAAAGT 429
QY 82 PheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGln 101
Db 430 TTTCCAGGCTTCATGGA-----TCAGAGATGTGGAAC 462
QY 102 AlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgPro---ProGly 120
Db 463 CCAACACACTGACCTCAGTGAAGACTGTTATATCTAAATGTATGATTCACAGCCTAA 522
QY 121 ThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyGlyPheGluIle 140
Db 523 CCAAAA-----AATGCCACTGATGATGATGATGATGATGATGATGATGATGATGAT 573
QY 141 GlySerProThrIlePheProAlaGlnMetValThrLysSerValLeuMetGlyLys 160

Db 574 GGAACATCATCTTTA-----CATGTTTATGAGCGCAG 606
Qy 161 -----HisIleHisValAlaValAsnTyrArgValAlaSerTrp 174
Db 607 TTTCTGGCTCGGTTGAAGAGTTATTGTAGTGTCAATGAAGTATAGGGTGGGTGCCCTA 666
Qy 175 GlyPheLeuAla-----GlyAspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeu 192
Db 667 GGAATCTTAGCTTGGCCAGGAAT-----CCTGAGGCTCCAGGGAACATGGGTTTA 717
Qy 193 LysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAsp 212
Db 718 TTTGATCAACAGTTGGCTCTTCAGTGGGTTCAAAAATAATAGACGCTTTGGTGGAAAT 777
Qy 213 ProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCyHisLeu 232
Db 778 CCTAAAGTGTAACTCTCTTTGGAGAAAGTGCAGGAGCAGCTTCAGTTAGCGCTGCATTG 837
Qy 233 IleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMet 252
Db 838 CTT---TCTCTCGGAAGCCATCA-----TTGTTCCAGAGCCCATTCGT 879
Qy 253 GlnSerGlyAlaMetValProSerAspProValAspGlyThrTyr-----GlyAsnGlu 270
Db 880 CAAGTGGATCATGAATGCTCTCTGGCGGTAACTCTTTATGAAGCTAGGAACAGA 939
Qy 271 IleTyrAspLeuPheValSerSerAlaGlyCysGly-----SerAlaSerAspLysLeu 288
Db 940 AGTTGAACTTA--GCTAAATTTGACTGGTCTCTAGAGAGAATGAGACTGAAATAATC 996
Qy 289 AlaCysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGly 308
Db 997 AAGTCTCTTAGAATAAAGATCCCAAGAAATCTT-----CTGAATGAAGCAATT 1047
Qy 309 PheLeuAlaTyrSerSer---LeuArgLeuSerTyrLeuProArgProAspGlyLysAsn 327
Db 1048 GTTGCCCTCTGGGACTCTTTGTCAGTAAACTTTGTCGACCGTGGATGGTATTTT 1107
Qy 328 IleThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValle 347
Db 1108 CTCACCTGACATGCCACACATATTACTTGAACCTGGACAATTTAAAMAAACCAGATTTG 1167
Qy 348 IleGlyAspGlnAsnAspGluGlyThrIlePhe-----GlyLeuSer 361
Db 1168 GTGGGTGTAAATAAGATGAAGGCAGACTTTTATGCTATGCTGCTCTGCTTCAGC 1227
Qy 362 SerLeuAsn-----ValThrThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPhe 379
Db 1228 AAGATAACATAGTATCATCACTAGAAAGAAATTTTCAGGAAGGTTTAAATAATTTT 1287
Qy 380 IleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyr-----394
Db 1288 CCAGGAGTGAGTGAGTTGGAAGGAATCCATCTTTTTCATTACACAGACTGGGTAGAT 1347
Qy 395 -----ProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsnAla 411
Db 1348 GATCAGAGACCTGAAACTACCGTGAGCGCTTGGGTGATGTTGTTGGGGATTAATTTTC 1407
Qy 412 LeuThrPro-----GlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIle 429
Db 1408 ATATGCCCTCGCTGGAGTTCACCAAGAAAGTCTCAGAAATGGGAATAATGCTTTTTC 1467
Qy 430 HisAlaArgArgTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSer 449
Db 1468 -----TACTATTTTGAACAC-----CGATCC 1488
Qy 450 LysGlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGln 469
Db 1489 TCCAAATCTCGTCGCCAGAAATGATGGAGTGATGATGGCTATGAAAT-----1539
Qy 470 AspTyrLeuLeuGly-----474

Db 1540 GAATTTGTCTTTGGTTTACCTCTGGAAAGAGAGATATTAATACACAAAGCCGAGAAATT 1599
Qy 475 ---SerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspPro 493
Db 1600 TTGAGTAGATCCCATAGTG-----AAMCGGTGGGCAAAATTTTGCAAAATATGGGAATCCA 1653
Qy 494 Asn---ThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGln 510
Db 1654 AATGAGACTCAGAACAAATAGCACAAAGCTGGCTGTCTTCAAAAGCACTGAACAA 1707
RESULT 8
US-10-324-466-39
; Sequence 39, Application US/10324466
; Publication No. US20040121970A1
; GENERAL INFORMATION:
; APPLICANT: Watkins, Jeffrey D.
; APPLICANT: Hancock, James D.
; TITLE OF INVENTION: Butyrylcholinesterase Variant
; TITLE OF INVENTION: Polypeptides with Increased Catalytic Efficiency and Methods
; FILE REFERENCE: P-IX 5555
; CURRENT APPLICATION NUMBER: US/10/324,466
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: US 10/032,233
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 2416
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Butyrylcholinesterase variant
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (214)...(1935)
US-10-324-466-39
Alignment Scores:
Pred. No.: 4,62e-44 Length: 2416
Score: 456.00 Matches: 160
Percent Similarity: 45.1% Conservative: 83
Best Local Similarity: 29.74% Mismatches: 185
Query Match: 15.92% Indels: 110
DB: 17 Gaps: 26
US-09-943-857-4 (1-547) x US-10-324-466-39 (1-2416)
Qy 22 GlyAspThrIleThrGlyLeuAsnAlaIleIleAsnGluAlaPheLeuGlyIleProPhe 41
Db 277 GGTGGCAGCGTAACA-----GCCTTTCTTGGAAATCCCTAT 312
Qy 42 AlaGluProProValGlyAsnLeuArgPheLysAspProValProTyrSerGlySerLeu 61
Db 313 GCACGCCACCTCTGGTAGACTTCGATTCAAAAGCCACAGTCTCTGACCAAGTGGTCT 372
Qy 62 AsnGlyGlnLysPheThrSerTyrGlyProSerCysMetGlnGlnAsnProGluGlyThr 81
Db 373 GATATTGGAAATGCCCAAAATATGCAAAATTTCTTGCTGT---CAGAACATAGATCAAGT 429
Qy 82 PheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGln 101
Db 430 TTTCCAGGCTTCCATGGA-----TCAGAGATGTGGAAC 462
Qy 102 AlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgPro---ProGly 120
Db 463 CCAAACTGACCTCAGTGAAGACTGTTATATCTAAATATATGATGATTCAGACACCTAAA 522
Qy 121 ThrIleAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyGlyPheGluIle 140
Db 523 CCAAAA-----AATGCCACTGATGATGATGATGATGATGATGATGATGATGATGAT 573
Qy 141 GlySerProThrIlePheProAlaGlnMetValThrLysSerValLeuMetGlyLys 160

```

Db 574 GGAACATCACTTTA-----CATGTTTATGATGGCAAG 606
Qy 161 -----HisIleHisValAlaValAsnTyrArgValAlaSerTrp 174
Db 607 TTCTCGGCTGGGTTGAAGAGTTATTCTAGTGTCAATGACTATAGGGTGGCCCTA 666
Qy 175 GlyPheLeuAla-----GlyAspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeu 192
Db 667 GGATTTCTTAGCTTTGCCAGGAAT-----CCTGAGGCTCCAGGGAACATGGGTTA 717
Qy 193 LysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyAsp 212
Db 718 TTGTATCAACAGTTGGCTCTTCATGGGTCTCAAAAAATATAGCAGCTTTGGTGGAAAT 777
Qy 213 ProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCysHisLeu 232
Db 778 CCTAAAAGTGTACTCTCTTTGGAGAAAGTCAGAGCAGCTTCAGTTAGCTGCATTG 837
Qy 233 IleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMet 252
Db 838 CTT---TCTCCTGAGCCATTCA-----TTGTTCCACGAGGCCATTCTG 879
Qy 253 GlnSerGlyAlaMetValProSerAspProValAspGlyThrTyr-----GlyAsnGlu 270
Db 880 CAAAGTGGATCCAGATGCTCCTTGGGCGGTAAACATCTCTTTATGAAGCTAGGAACAGA 939
Qy 271 IleTyrAspLeuPheValSerSerAlaGlyCysGly-----SerAlaSerAspLysLeu 288
Db 940 ACCTTGAACCTTA---GCTAAATGACTGGTGTCTAGAGAGATGAGCTGAAATATC 996
Qy 289 AlaCysLeuArgSerAlaSerAspThrLeuLeuAspAlaThrAsnAsnThrProGly 308
Db 997 AAGTGTCTTAGAAATAAGATCCCAAGAAATCTT-----CTGAATGAAGCATTT 1047
Qy 309 PheLeuAlaTyrSerSer---LeuArgLeuSerTyrLeuProArgProAspGlyLysAsn 327
Db 1048 GTTGTCCTCATGGGACTCTTTGTGAGTAAACTTTGGTCCGACCGGTGATGGTATTT 1107
Qy 328 IleThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIle 347
Db 1108 CTCACCTGACATGCCAGACATATTACTTGAACCTTGGCAATTTAAAAACCCAGATTTTG 1167
Qy 348 IleGlyAspGlnAsnAspGluGlyThrIlePhe-----GlyLeuSer 361
Db 1168 GTGGGTGTTAATAAAGATGAAGGACAGACTTTTATTAGTCTATGCTCTCTGCTTCAGC 1227
Qy 362 SerLeuAsn-----ValThrThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPhe 379
Db 1228 AAAGATACAATAGTATCATATACTAGAAAGAAATTTTCAGGAAGGTTTAAAAATATTTT 1287
Qy 380 IleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyr-----394
Db 1288 CCAGGAGTGAAGTGTGGTGAAGGAATCCATCTTTTTCATTACACAGCTGGGTAGAT 1347
Qy 395 -----ProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsnAla 411
Db 1348 GATCAGAGACCTGAAACACTACCGTGGGCTTGGGTGATGTTGTTGGGATTAATAATTC 1407
Qy 412 LeuThrPro-----GlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIle 429
Db 1408 ATATGCCCTGCGCTTGAGTTTCCACCAAGAGTCTCAGAAATGGGGAATAATGCTTTTTC 1467
Qy 430 HisAlaArgTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSer 449
Db 1468 -----TACTATTTTGAACAC-----CGATCC 1488
Qy 450 LysGlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGln 469
Db 1489 TCCAAACTTCCGTGGCCGAATGGATGGGAGTGATGCGTATGAATTT-----1539
Qy 470 AspTyrLeuLeuGly-----474

```

```

Db 1540 GAATTTCTGTTGGTTTACCTCTGGAAGAAGAGATAAATTACACAAAAGCCGAGAAATT 1599
Qy 475 ---SerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspPro 493
Db 1600 TTGAGTAGATCCATAGTGT-----AAACGGTGGGCAAAATTTTGCAAAATATGGGAATCCA 1653
Qy 494 Asn---ThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGln 510
Db 1654 AATGAGACTCAGAACATAGCACAAGCTGGCTCTCTTCAAAGCACTGAACAA 1707

RESULT 9
US-10-413-432-39
; Sequence 39, Application US/10413432
; Publication No. US20040120939A1
; GENERAL INFORMATION:
; APPLICANT: Watkins, Jeffrey D.
; APPLICANT: Hancock, James D.
; TITLE OF INVENTION: Butyrylcholinesterase Variant
; TITLE OF INVENTION: Polypeptides with Increased Catalytic Efficiency and Methods
; FILE REFERENCE: P-IX 5510
; CURRENT APPLICATION NUMBER: US/10/413,432
; CURRENT FILING DATE: 2003-04-11
; PRIOR APPLICATION NUMBER: US 10/324,466
; PRIOR FILING DATE: 2002-12-20
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 2416
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Butyrylcholinesterase variant
; NAME/KEY: CDS
; LOCATION: (214) ... (1935)
US-10-413-432-39

Alignment Scores:
Pred. No.: 4,62e-44 Length: 2416
Score: 456.00 Matches: 160
Percent Similarity: 45.17% Conservative: 83
Best Local Similarity: 29.74% Mismatches: 185
Query Match: 15,92% Indels: 110
DB: 17 Gaps: 26

US-09-943-857-4 (1-547) x US-10-413-432-39 (1-2416)
Qy 22 GlyAspThrIleThrGlyLeuAsnAlaIleIleAsnGluAlaPheLeuGlyIleProPhe 41
Db 277 GGTGGCACGGTAACA-----GCCTTTCTTGGAAATTCCTAT 312
Qy 42 AlaGluProProValGlyAsnLeuArgPheLysAspProValProTyrSerGlySerLeu 61
Db 313 GCACAGCCACCTCTTGTAGACTTCGATTCAAAAAGCCACAGCTCTCTGACCAAGTGTCT 372
Qy 62 AsnGlyGlnLysPheThrSerTyrGlyProSerCysMetGlnGlnAsnProGluGlyThr 81
Db 373 GATATTTGGAAATGCCACAAAATATCAAAATTCCTCTGT---CAGAACATATGCAAGT 429
Qy 82 PheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGln 101
Db 430 TTTCCAGGCTTCCATGGA-----TCAGAGATGTGGAAC 462
Qy 102 AlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgPro---ProGly 120
Db 463 CCAAACTGACCTCAGTGAAGACTGTTTATATCTAAATGATGATTCACACACCTAAA 522
Qy 121 ThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyGlyPheGluIle 140
Db 523 CCAAAA-----AATGCCACTGTATTGATGATTTATGTTGTTGTTTCAAACT 573
Qy 141 GlySerProThrIlePheProAlaGlnMetValThrLysSerValLeuMetGlyLys 160

```

```

Db 574 GGAACATCATCTTTA-----CATGTTTATGATGCAAG 606
Qy 161 -----HisIleIleHisValAlaValAsnTyrArgValAlaSerTrp 174
Db 607 TTTCTGCTCGGTTGAAGAGTATTGTAGTGTCAATGAATATATAGGTGGTGGCTTA 666
Qy 175 GlyPheLeuAla-----GlyAspAspIleIleValAlaGluGlySerGlyAsnAlaGlyLeu 192
Db 667 GGATTCCTAGCTTTGCCAGAAAT-----CCTGAGGCTCCAGGGAACATCGGTTTA 717
Qy 193 LysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAsn 212
Db 718 TTTGATCAACAGTTGGCTCTTCATGGTTCAGGTTCAAAAATAATAGCAGCTTTGGTGAAT 777
Qy 213 ProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCysHisLeu 232
Db 778 CTTAAAGTGTAACTCTCTTTGGAGAAAGTGCAGAGCAGCTTCAGTTAGCTGCCTCATTTG 837
Qy 233 IleTrpAsnAspGlyAspAsnThrTyrIleGlyLysProLeuPheArgAlaGlyIleMet 252
Db 838 CTT---TCTCTCGAAGCCATCA-----TTGTTCCAGCAGAGCCATTCG 879
Qy 253 GlnSerGlyAlaMetValProSerAspProValAspGlyThrTyr-----GlyAsnGlu 270
Db 880 CAAGTGGATCCATGAATGCTCTTGGCGGTAAATCTCTTTATGAAGCTAGGAACAGA 939
Qy 271 IleTyrAspLeuPheValSerAlaGlyCysGly-----SerAlaSerAspLysLeu 288
Db 940 ACGTTGAACCTTA---GCTAAATGACTGTGCTCTAGAGAGAATAGAGCTGAATAATC 996
Qy 289 AlaCysLeuArgSerAlaSerAspThrLeuLeuAspAlaThrAsnAsnThrProGly 308
Db 997 AAGTGTCTTAGAAATAAGATCCCAAGAAATCTT-----CTGAATGACCATTT 1047
Qy 309 PheLeuAlaTyrSerSer---LeuArgLeuSerTyrLeuProArgProAspGlyLysAsn 327
Db 1048 GTTGTCCTATAGGACTCTTGTTCAGTAAACTTTGTCGACCGTGGATGATTTT 1107
Qy 328 IleThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIle 347
Db 1108 CTCACCTGACATCCAGACATATTACTTGAACCTTGACAAATTTAAAAAACCAGATTTG 1167
Qy 348 IleGlyAspGlnAsnAspGluGlyThrIlePhe-----GlyLeuSer 361
Db 1168 GTGGGTGTATAAAGATGAAGGACAGCTTTTATGCTATGGTCTCTCGCTTCAGC 1227
Qy 362 SerLeuAsn-----ValThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPhe 379
Db 1228 AAAGATNACAATAGTATCATCACTAGAAAAGAAATTCAGGAAGGTTTAAAAATATTTT 1287
Qy 380 IleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyr----- 394
Db 1288 CCAGGAGTGAGTGAGTTTGGAAAGGAATCCATCTTTTCATTACACAGCTGGGTAGAT 1347
Qy 395 -----ProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsnAla 411
Db 1348 GATCAGACCTGAAACACTACGCTGAGCGCTGGGTGATGTTGTTGGGATTAATAATTC 1407
Qy 412 LeuThrPro-----GlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIle 429
Db 1408 ATATGCCCTGCTCGGAGTTCCACAGAAGTTCTCAGAAATGGGAAATAATGCTTTTTC 1467
Qy 430 HisAlaArgTyrPheLeuAsnHisPheGlnGlyGlyThrLysTyrSerPheLeuSer 449
Db 1468 -----TACTATTTTGAACAC-----CGATCC 1488
Qy 450 LysGlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGln 469
Db 1489 TCCAAACTTCCGTGGCCAGATGGATGGAGTGCATGCTATGAAAT----- 1539
Qy 470 AspTyrLeuLeuGly----- 474

```

```

Db 1540 GAATTTGTCTTTGGTTTACCTCTGGAAAGAGAGATATTACACAAAGCCGAGAAATT 1599
Qy 475 ---SerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspPro 493
Db 1600 TTGAGTAGATCCATAGTG-----AAACGGTGGCAAATTTTGCATAATATGGGAATCCA 1653
Qy 494 Asn---ThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGln 510
Db 1654 AATGAGACTCAGAACCAATAGCACAAAGCTGCCTGTCTTCAAAAGCACTGAACAA 1707

RESULT 10
US-09-748-739A-7
; Sequence 7, Application US/09748739A
; Patent No. US2002011948A1
; GENERAL INFORMATION:
; APPLICANT: Lockridge, Oksana
; TITLE OF INVENTION: Butyrylcholinesterase Variants and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: P-IX 4143
; CURRENT FILING DATE: 2000-12-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 2416
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Human Butyrylcholinesterase variant
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (214)...(1935)
US-09-748-739A-7

Alignment Scores:
Pred. No.: 6,1e-44 Length: 2416
Score: 455.00 Matches: 160
Percent Similarity: 45.17% Conservative: 83
Best Local Similarity: 29.74% Mismatches: 185
Query Match: 15.89% Indels: 110
DB: 9 Gaps: 26

US-09-943-857-4 (1-547) x US-09-748-739A-7 (1-2416)
Qy 22 GlyAspThrIleThrGlyLeuAsnAlaIleIleAsnGluAlaPheLeuGlyIleProPhe 41
Db 277 GTGGCACCGGTAACA-----GCCTTTCTTGAATTCCTAT 312
Qy 42 AlaGluProValGlyAsnLeuArgPheLysAspProValProTyrSerGlySerLeu 61
Db 313 GCACAGCCACCTCTTGTAGACTTCGATTCAAAAAGCCACAGCTCTCTGACCAAGTGGTCT 372
Qy 62 AsnGlyGlnLysPheThrSerTyrGlyProSerCysMetGlnGlnAsnProGluGlyThr 81
Db 373 GATATTGGATGCGCACAAATATGCAAAATCTTCTGTCT---CAGAACATAGATCAAACT 429
Qy 82 PheGluGlnAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGln 101
Db 430 TTTCCAGGCTTCATGGA-----TCAGAGATGTGAAC 462
Qy 102 AlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgPro---ProGly 120
Db 463 CCAAAACACTGACCTCAGTGAAGACTGTTTATATCTAAATGTATGATTCAGCACCTAAA 522
Qy 121 ThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyGlyGlyPheGluIle 140
Db 523 CCAAAA-----AATGCCACTGATTGATGATGATTTATGTTGGTGGTGGTTTCAAACT 573
Qy 141 GlySerProThrIlePheProAlaGlnMetValThrLysSerValLeuMetGlyLys 160
Db 574 GGAACATCATCTTTA-----CATGTTTATGATGCAAG 606

```

```

QY 161 -----HisIleIleHisValalaValaAsnTyrArgValalaSerTrp 174
Db 607 TTTCTGGCTCGGGTTGAAGAGATTATTGTAGTCAATGAACACTATAGGGTGGCCTA 666
QY 175 GlyPheLeuAla-----GlyAspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeu 192
Db 667 GGATTCCTTAGCTTGGCAGGAAAT-----CCTGAGGCTCCAGGGGAACATGGGTTA 717
QY 193 LysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAsp 212
Db 718 TTTGATCAACAGTTGGCTCTTCAGTGGTTCAAAAAATATAGCAGCTTTGGTGGAAAT 777
QY 213 ProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCyHisLeu 232
Db 778 CCTAAAGTGTAACTCTCTTTGGAGAAAGTCAGAGCAGCTTCAGTTAGCTGCATTG 837
QY 233 IleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMet 252
Db 838 CTT---TCTCTGGAAGCCATTCA-----TTGTTCCACGAGCCATTCTG 879
QY 253 GlnSerGlyAlaMetValProSerAspProValAspGlyThrTyr-----GlyAsnGlu 270
Db 880 CAAAGTGCATCTTAATGCTCTTGGCGGTAACTCTCTTATGAGCTAGAGTAAGACAGA 939
QY 271 IleTyrAspLeuPheValSerAlaGlyCysGly-----SerAlaSerAspLysLeu 288
Db 940 ACGTGTGAACCTA---GCTAAATTGACTGTGTCTAGAGAGATGAGACTGAATAATC 996
QY 289 AlaCysLeuArgSerAlaSerAspThrLeuLeuAspAlaThrAsnAsnThrProGly 308
Db 997 AAGTGTCTTGAATAAAGATCCCAAGAAATCTT-----CTGAATGAAGCAATTT 1047
QY 309 PheLeuAlaTyr---SerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAsn 327
Db 1048 GTGTGCCCCATGCGGACTGTGTGTCAGTAACTTTGGTCGACCGGTGAGTGTATTT 1107
QY 328 IleThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIle 347
Db 1108 CTCACGTACATGCCAGACATATTACTTGAACCTTGGACAAATTAAMAAAACCCAGATTTG 1167
QY 348 IleGlyAspGlnAsnAspGluGlyThrIlePhe-----GlyLeuSer 361
Db 1168 GTGGGTGTTAATAAGATGAGGAGCAGCTTTTCTAGTCTATGGTGTCTCTGCTTCAGC 1227
QY 362 SerLeuAsn-----ValThrThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPhe 379
Db 1228 AAAGATAACAATAGTATCATCACTAGAAAAGATTTTCAGGAAGGTTTAAATAATTTTT 1287
QY 380 IleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyr-----394
Db 1288 CCAGGAGTGAAGTGTGGTGGAAAGGAATCCATCCTTTTTCATTACACAGACTGGTAGAT 1347
QY 395 -----ProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsnAla 411
Db 1348 GATCAGACCTGGAACACTACCGTAGGCGCTGGGTGATGTTGTTGGGGATTATAATTTTC 1407
QY 412 LeuThrPro-----GlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIle 429
Db 1408 ATATGCCCTGCTTGGAGTTCACCAAGAGTCTCAGAAATGGGGAATAATGCTTTTTC 1467
QY 430 HisAlaArgArgTyrPheLeuAsnHisPheGlnGlyGlyThrLysTyrSerPheLeuSer 449
Db 1468 -----TACTATTTTGAACAC-----CGATCC 1488
QY 450 LysGlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGln 469
Db 1489 TCCAAACTTCCGTGGCCAGAAATGGATGGGAGTATGATGGCTATGAAAT-----1539
QY 470 AspTyrLeuLeuGly-----474
Db 1540 GAATTGTCTTTGGTTTACCTCTGGAAGAGAGATAATTACAAAAACCCGAGGAATTT 1599
QY 475 ---SerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspPro 493

```

```

Db 1600 TTGAGTAGATCCATAGTG-----AAACGGTGGGCAATTTTGCATAATATGGGAATCCA 1653
QY 494 Asn---ThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGln 510
Db 1654 AATGAGACTCAGAACAAATAGCACAAAGCTGGCTGTCTTCAAAGACACTGAACAA 1707
RESULT 11
US-09-810-861B-5
; Sequence 5, Application US/09810861B
; Patent No. US20020162140A1
; GENERAL INFORMATION:
; APPLICANT: Mor, Tsafir S.
; APPLICANT: Soreq, Hermona
; APPLICANT: Arentzen, Charles J.
; APPLICANT: Maeson, Hugh S.
; TITLE OF INVENTION: EXPRESSION OF RECOMBINANT HUMAN ACETYLCHOLINESTERASE IN
; FILE REFERENCE: BTI-45
; CURRENT APPLICATION NUMBER: US/09/810,861B
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/190,440
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 5
; LENGTH: 1725
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: human acetylcholinesterase gene optimized for
; OTHER INFORMATION: expression in plants
US-09-810-861B-5
Alignment Scores:
Pred. No.: 8,21e-44 Length: 1725
Score: 452.00 Matches: 156
Percent Similarity: 44.59% Conservative: 83
Best Local Similarity: 29.10% Mismatches: 179
Query Match: 15.78% Indels: 118
DB: Gaps: 26
US-09-943-857-4 (1-547) x US-09-810-861B-5 (1-1725)
QY 35 AlaPheLeuGlyIleProPheAlaGluProValGlyAsnLeuArgPheLysAspPro 54
Db 184 GCTTCTCGGCATCCCTTTGCGAGAGCCACCCATGGGACCCCGTCGCTTCTGCCACCG 243
QY 55 -----ValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyPro 71
Db 244 GAGCCCAACGACGCTTGGTCAGGGGTGTAGACGCTACCACTTCAGAGTGTCTC-----297
QY 72 SerCysMetGln-----GlnAsnProGluGlyThrPheGluGluAsn 85
Db 298 ---TGCTACCATATGTGGACACCCCTATACCCAGGTTTGGAGGACCCAGAGATGTGAAC 354
QY 86 LeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuPro 105
Db 355 CCCAACCCGTGAGCTG-----369
QY 106 GlnSerGluAspCysLeuThrIleAsnValValArgProGlyThrLysAlaGlyAla 125
Db 370 ---AGCAGGAGCTGCTGTACCTCAAC---GTGTGGACACCACTACCCCGGCTCATATCC 423
QY 126 AsnLeuProValMetLeuTrpIlePheGlyGlyGlyPheGluIleGlySerProThrIle 145
Db 424 CCCACCCCTGCTCTGCTGTGATCTATGGGGTGGCTTCTACAGTGGGGGCTCTCTCTTG 483
QY 146 PheProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIleHis---164
Db 484 -----GACGTGTACGATGGCGCTTCTTGTGTACAGGCC 516

```



```

165 QY -----ValAlaValAsnTyrArgValAlaSerTyrGlyPheLeuAlaGly 179
166 QY |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 Db GAGAGGACTGTGGTGTGCATCACTACCGGTGGGAGCCCTTGGCTTCTTGGCCCTG 576
168 QY 180 AspAspIleLysAlaGluGlySerGlyAsnAlaGlyLysAspGlnArgLeuGlyMet 199
169 Db 577 CCGGGGAGCGA---GAGGCCCCGGGCAATGTGGTCTCTGGATCAGAGGCTGGCCCTG 633
170 QY 200 GlnTrpValAlaAspAsnIleAlaGlyPheGlyAspProSerLysValThrIlePhe 219
171 Db 634 CAGTGGGTGAGGAGAGAGCTGGCAGCCTTCGGGGGTGACCCGACATCAGTGCAGCCTGTT 693
172 QY 220 GlyLysSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAspGlyAspAsn 239
173 Db 694 GGGGAGAGCGGGAGCGGCTCGGTGGGCAATGCACCTGCTG-----735
174 QY 240 ThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValPro 259
175 Db 736 TCCCGCCGAGCGGGGCTGTTCCACAGGCGCGTGTGCAGAGCGGTGCC-----CCC 789
176 QY 260 SerAspProValAspGlyThrTyrGly-----AsnGluIle 271
177 Db 790 AATGGACCCCTGG---GCCACCGTGGGCAATGGAGAGCGCCCTGCGAGGCGCCACGAGCTG 846
178 QY 272 TyrAspLeuPheValSerAlaGlyCys-----GlySerAlaSer 285
179 Db 847 GCCCACCTT-----GTGGCTGTCTCTCCAGGCGGCACTGGTGGGATGACACA 894
180 QY 286 AspLysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsn 305
181 Db 895 GAGCTGTAGCTGCTCGGACACGACGACGAGCGGAGTCTGGT-----AACAC 945
182 QY 306 ThrProGlyPheLeuAlaTyrSerSerLeu---ArgLeuSerTyrLeuProArgPro 324
183 Db 946 GAATGGCAGTGTCTCCCTCAGAAAGCGTCTTCCGGTCTCTCTCGTGGTGTGTAGAT 1005
184 QY 325 GlyLysAsnIleThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerVal 344
185 Db 1006 GGAGACTTCTCAGTGAACACCCAGAGGCGCTCATCAGCGGGAGACTTCCACGCGCTG 1065
186 QY 345 ProValIleIleGlyAspGlnAsnAspGluGlyThrIlePhe-----358
187 Db 1066 CAGGTGCTGTGGTGTGTGAAGGATGAGGCTCGTATTTCTGGTTTACGGGCCCCA 1125
188 QY 359 GlyLeuSerSerLeuAsn-----ValThrThrAsnAlaGlnAlaArgAlaTyrPheLys 376
189 Db 1126 GGCCTTCAGCAAGACAAACAGTCTCTCATCAGCGGGCGGAGTTCCTGGCGGGGTGG 1185
190 QY 377 GlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyr-----394
191 Db 1186 GTGGGGTTCCTCAGGTAAAGTACCTGGCGGCGGAGGCTGTGGTCTGTCATACAGAC 1245
192 QY 395 -----ProGlnAsp-----IleThrGlnGlySerProPheAspThrGlyVal 408
193 Db 1246 TGGCTGCATCCCGAGGACCCGCGCAGCTGTGAGGAGGCGCTGAGCGATGTGTGGCGAG 1305
194 QY 409 LeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPhe 428
195 Db 1306 CCAATGTGTGTGGTCC-----GTGGCCAGTGTGGTGGGAGCTGGTGGTGGC 1353
196 QY 429 IleHisAlaArg-----ArgTyrPheLeuAsnHisPheGlnGly-----441
197 Db 1354 CAGGTGCGCGGTCTACGCTTGTGAACACCGTCTTCCAGGCTCTCTGGGCC 1413
198 QY 442 -----GlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeuProIle---457
199 Db 1414 CTGTGATGGGGTGGCCACGCTACGATCGATTCATCTTGGGATCCCGCTGGAC 1473
200 QY 458 ---MetGlyThrPheHisAlaAsnAspIleValTyrGlnAspTyrLeuLeuGlySerGly 476
201 Db 1474 CCCTTCGAAGACTACGCGGAGGAGAAATCTTCGCCAGGAGCTGATG-----1524
202 QY 477 SerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAla 496
203 QY |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
204 Db 1525 -----CGAFACTGGGCCAACTTTGGCCGACACAGGGATCCCAATGAGCCC 1569
205 QY 497 -----GlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGln 510
206 Db 1570 CGAGACCCCAAGCCCCCAATGCCCCCGTACACGCGGGGGTCTCAG 1617
207 RESULT 12
208 US-10-032-233-33
209 ; Sequence 33, Application US/10032233
210 ; Publication No. US20030153062A1
211 ; GENERAL INFORMATION:
212 ; APPLICANT: Watkins, Jeffrey D.
213 ; APPLICANT: Hancock, James D.
214 ; TITLE OF INVENTION: Butyrylcholinesterase Variants with
215 ; TITLE OF INVENTION: Increased Catalytic Efficiency and Methods of Use
216 ; FILE REFERENCE: P-IX 4642
217 ; CURRENT APPLICATION NUMBER: US/10/032,233
218 ; CURRENT FILING DATE: 2001-12-20
219 ; NUMBER OF SEQ ID NOS: 50
220 ; SOFTWARE: FastSeq for Windows Version 4.0
221 ; SEQ ID NO 33
222 ; LENGTH: 2416
223 ; TYPE: DNA
224 ; ORGANISM: Artificial Sequence
225 ; FEATURE:
226 ; OTHER INFORMATION: Butyrylcholinesterase variant
227 ; NAME/KEY: CDS
228 ; LOCATION: (214)....(1935)
229 US-10-032-233-33
230 Alignment Scores:
231 Pred. No.: 1,41e-43 Length: 2416
232 Score: 452.00 Matches: 161
233 Percent Similarity: 45.37% Conservative: 84
234 Best Local Similarity: 29.81% Mismatches: 181
235 Query Match: 15.78% Indels: 114
236 DB: 15 Gaps: 28
237 US-09-943-857-4 (1-547) x US-10-032-233-33 (1-2416)
238 QY 22 GlyAspThrIleThrGlyLeuAsnAlaIleIleAsnGluAlaPheLeuGlyIleProPhe 41
239 Db 277 GGTGGCAGCGGTAAACA-----GCCTTCTTGGATTTCCCTAT 312
240 QY 42 AlaGluProProValGlyAsnLeuArgPheLysAspProValProTyrSerGlySerLeu 51
241 Db 313 GCACAGCCACCTCTTGGTAGACTTCGATTCAAAAAGCCACAGTCTCTGACCAAGTGTCT 372
242 QY 62 AsnGlyGlnLysPheThrSerTyrGlyProSerCysMetGlnGlnAsnProGluGlyThr 81
243 Db 373 GATATTGGAAATGCCACAAATATGCAATTTCTGCTGT---CAGAACATAGATCAAGT 429
244 QY 82 PheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGln 101
245 Db 430 TTTCAGGCTTCCATGGA-----TCAGATGTGGAAC 462
246 QY 102 AlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgPro---ProGly 120
247 Db 463 CCAACACATGACCTCAGTCAAGACTGTTTATATCATTAATGTATGATTCAGCACCCTAA 522
248 QY 121 ThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyGlyGlyPheGluIle 140
249 Db 523 CCAAAA-----AATGCCACTGTATTGATGATGATTTATGTTGGTGGTGGTTTCAACT 573
250 QY 141 GlySerProThrIlePheProProAlaGlnMetValThrLysSerValLeuMetGlyLys 160
251 Db 574 GGAACATCATCTTA-----CATGTTTATGATGGAAC 606
252 QY 161 -----HisIleIleHisValAlaValAsnTyrArgValAlaSerTyr 174
253 Db 607 TTCTCGCTCGGTGTAAGAGATTATTGTAGTGTCAATGAATATAGGGGTGGGTGCCCTA 666

```

```

QY 175 GlyPheLeuAla-----GlyAspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeu 192
DB 567 GGATTCCTTAGCTTCCAGCAAAAT-----CCTGAGGCTCCAGGCAACATGGGTTTA 717

QY 193 LysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAsp 212
DB 718 TTTCATCAACAGTTGGCTCTTCAGTGGGTTCAAAAAATATAGCAGCCTTTGGTGGAAAT 777

QY 213 ProSerLysValThrIlePheGlyGlySerAlaGlySerMetSerValLeuLysCysHisLeu 232
DB 778 CCTAAAAGTGTAACTCTCTTTGGAGAAAGTGCAGGAGCAGCTTCAGTTAGCTGCATTTG 837

QY 233 IleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyLeuMet 252
DB 838 CTT----TCTCTGGAAGCCATCA-----TTGTTCCAGCAGCCATCTGT 879

QY 253 GlnSerGlyAlaMetValProSerAspPro-----ValAspGlyThrTyr-----Gly 268
DB 880 CAAAGTGGATCC-----CCGAATGCTCTTGGCGGTAACATCTTTATGAAAGCTAGG 933

QY 269 AsnGluIleTyrAspLeuPheValSerAlaGlyCysGly-----SerAlaSerAsp 286
DB 934 AACAGAACGTTGAACCTTA---GCTAAATTGACTGTTGCTCTAGAGAGAAATGACACTGAA 990

QY 287 LysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThr 306
DB 991 ATAACTCAAGTGTCTTGAATAAAGATCCCAAGAAATCTT-----CTGAATGAA 1041

QY 307 ProGlyPheLeuAlaTyrSerSer---LeuArgLeuSerTyrLeuProArgProAspGly 325
DB 1042 GCATTGTGTCCTCCATGGAGCTCTCTTGTGCTAAACCTTTGTCGCGCGGTGGATGGT 1101

QY 326 LysAsnIleThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValPro 345
DB 1102 GATTTCTCTACTGACATGCCAGACATATTACTTGAACCTTGACAAATTTAAAAAACCCAG 1161

QY 346 ValIleIleGlyAspGlnAsnAspGluGlyThrIlePhe-----Gly 359
DB 1162 ATTTTGGTGGTGTAAATAAAGATGAAGGAGCAGCTTTTTTACTGTATGGTCTCTCGGC 1221

QY 360 LeuSerSerLeuAsn-----ValThrThrAsnAlaGlnAlaArgAlaTyrPheLysGln 377
DB 1222 TTCAGCAAGATCAACATAGTATCATCACTAGAAAAGAAATTTCCAGGAAGGTTTAAAAATA 1281

QY 378 SerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaIatyr-----394
DB 1282 TTTTTCAGGAGTGTAGTGGTTGGAAGGAATCCATCTCTTTTTCATTACACAGACTGG 1341

QY 395 -----ProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeu 409
DB 1342 GTAGATGATCAGAGACCTGAAACTACCGTGGAGCCCTTGGGTGATGTTGTTGGGGATTAT 1401

QY 410 AsnAlaLeuThrPro-----GlnPheLysArgIleSerAlaValLeuGlyAspLeuAla 427
DB 1402 AATTCATATGCCCTGCTTGGAGTTCCACAGAAGTCTCCGAATGGGGAATAATATGCC 1461

QY 428 PheIleHisAlaArgTyrPheLeuAsnHisPheGlnGlyThrIlystYrSerPhe 447
DB 1462 TTTTTC-----TACTATTTTGAACAC-----1482

QY 448 LeuSerLysGlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleVal 467
DB 1483 CGATCTCCCAACTCTCCGCGCAGAAATGGATGGAGTGATGATGATGATGATGATGATGATGAT 1539

QY 468 TrpGlnAspTyrLeuLeuGly-----1540
DB 1540 -----GAATTTGCTTTGGTTTACCTCTGGAAGAGAGATAATTACAAAAAGCCGAG 1593

QY 475 -----SerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeu 491
DB 1594 GAAATTTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1647

QY 492 AspProAsn---ThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerGln 510

```

```

DB 1648 AATCCAAATGAGACTCAGAACATAGCACAGCTGGCTGTCTTCAAAGGCACTGACAA 1707

RESULT 13
US-10-324-466-33
; Sequence 33, Application US/10324466
; Publication No. US20040121970A1
; GENERAL INFORMATION:
; APPLICANT: Watkins, Jeffrey D.
; APPLICANT: Hancock, James D.
; TITLE OF INVENTION: Butyrylcholinesterase Variant
; TITLE OF INVENTION: Polypeptides with Increased Catalytic Efficiency and Methods
; FILE REFERENCE: P-IX 5555
; CURRENT APPLICATION NUMBER: US/10/324,466
; CURRENT FILING DATE: 2002-12-20
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 2416
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Butyrylcholinesterase variant
; NAME/KEY: CDS
; LOCATION: (214) ... (1935)
US-10-324-466-33

```

Alignment Scores:

```

Pred. No.: 1,416-43 Length: 2416
Score: 452.00 Matches: 161
Percent Similarity: 45.37% Conservative: 84
Best Local Similarity: 29.81% Mismatches: 181
Query Match: 15.78% Indels: 114
DB: 17 Gaps: 28

```

US-09-943-857-4 (1-547) x US-10-324-466-33 (1-2416)

```

QY 22 GlyAspThrIleThrGlyLeuAsnAlaIleAsnGluAlaPheLeuGlyIleProPhe 41
DB 277 GGTGGCAGGTAAACA-----GCCTTCTTGAATTCCTAT 312

QY 42 AlaGluProValGlyAsnLeuArgPheLysAspProValProTyrSerGlySerLeu 61
DB 313 GCACGCCACCTCTTGGTAGACTTCGATTCAAAAAGCCACAGTCTCTGACCAAGTGTCT 372

QY 62 AsnGlyGlnLysPheThrSerTyrGlyProSerCysMetGlnGlnAsnProGlyThr 81
DB 373 GATATTGGAAATGCCCAAAATATGCAAAATCTTGTCTGT---CAGAACATAGATCAAAAGT 429

QY 82 PheGluGlnAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGln 101
DB 430 TTTCCAGGCTTCCATGGA-----TCAGAGATGTGGAAC 462

QY 102 AlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgPro---ProGly 120
DB 463 CCAAAACACTGACCTCAGTGAAGACTTTATATCTAAATGATGATGATGATGATGATGATGAT 522

QY 121 ThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyGlyGlyPheGluIle 140
DB 523 CCAAAA-----AATGCCACTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 573

QY 141 GlySerProThrIlePheProAlaGlnMetValThrLysSerValLeuMetGlyLys 160
DB 574 GGAACATCATCTTA-----CATGTTTATGATGCGCAG 606

QY 161 -----HisIleIleHisValAlaValAsnTyrArgValAlaSerTrp 174
DB 607 TTTCTGGCTCGGTTGAAGAGATTATTAGTGTCAATGAACATATAGGTTGGTGGCTCCTTA 666

```


QY 175 GlyPheLeuAala-----GlyAspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeu 192
 DB 667 GATTCCTTAGCTTCCAGGAAT-----CCTGAGGCTCCAGGAACATGGGTTTA 717
 QY 193 LysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyAsp 212
 DB 718 TTTCATCAACAGTTGGCTCTTCAGTGGGTTCAAAAAATATAGCAGCCTTTGGTGAAT 777
 QY 213 ProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuGlyHisLeu 232
 DB 778 CCTAAAGTGTAACCTCTTTGGAGAAAGTGCAGGAGCAGCTTCAGTTAGCCTGCATTTG 837
 QY 233 IleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMet 252
 DB 838 CTT---TCTCTGGAAGCCATCA-----TTGTTTCCAGGAGCCATTCG 879
 QY 253 GlnSerGlyAlaMetValProSerAspPro-----ValAspGlyThrTyr-----Gly 268
 DB 880 CAAGTGGATCC-----CCGAATGCTCTTGGGGGGTAAACATCTCTTATGAAGCTAGG 933
 QY 269 AsnGluIleTyrAspLeuPheValSerSerAlaGlyCysGly-----SerAlaSerAsp 286
 DB 934 AACAGAAGCTTGAACCTTA---GCTAAATTGACTGTTGCTCTAGAGAGATGAGACTGAA 990
 QY 287 LysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThr 306
 DB 991 ATATCAAGTGTCTTAGAATAAAGATCCCAAGAAATTCCTT-----CTGAATGAA 1041
 QY 307 ProGlyPheLeuAlaTyrSerSer---LeuArgLeuSerTyrLeuProArgProAspGly 325
 DB 1042 GCATTTGTTGCTCCCTATGGACTCCTTTGTCAGTAACCTTTGTCGACCGCTGGATGGT 1101
 QY 326 LysAsnIleThrAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValPro 345
 DB 1102 GATTTTCTCACTGACATGCAGACATATATTCTGAACTTGACATTTTAAAAACCCAG 1161
 QY 346 ValIleIleGlyAspGlnAsnAspGluGlyThrIlePhe-----Gly 359
 DB 1162 ATTTTGGTGGTGTATAAAGATGAAGGACAGCTTTTATGTCATGCTGCTCCTGGC 1221
 QY 360 LeuSerSerLeuAsn-----ValThrAsnAlaGlnAlaArgAlaTyrPheLysGln 377
 DB 1222 TTCAGCAAGATACATAGATATCATCACTAGAAAGATTTTCAGGAAGTTTAAAAATA 1281
 QY 378 SerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyr-----394
 DB 1282 TTTTTCAGAGTGTAGTGGTGGAGAGAAATCCATCCTTTTTCATTACAGACTGG 1341
 QY 395 -----ProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeu 409
 DB 1342 GTAGATGATCAGAGACTGAAACTACCTGAGGCTTGGGTGATGTTGTTGGGATAT 1401
 QY 410 AsnAlaLeuThrPro-----GlnPheLysArgIleSerAlaValLeuGlyAspLeuAla 427
 DB 1402 AATTTTCATATGCTCCTGCTGGAGTTTCCACCAAGTTCCTCAGATGGGGAATAATGCC 1461
 QY 428 PheIleHisAlaArgTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPhe 447
 DB 1462 TTTTTC-----TACTATTGTAACAC-----1482
 QY 448 LeuSerLysGlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleVal 467
 DB 1483 CGATCTCCAAACTCCGTTGGCAGAAATGGATGGAGTATGATGCTATGAAAT---1539
 QY 468 TrpGlnAspTyrLeuLeuGly-----474
 DB 1540 -----GAATTTGCTTTGGTTTACCTCTGGAAGAAAGAGATAATATACAAAAAGCCGAG 1593
 QY 475 -----SerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeu 491
 DB 1594 GAAATTTTCAGTAGATCCATAGT-----AAACGTTGGGCAAAATTTGCCAAATATGGG 1647
 QY 492 AspProAsn---ThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerGln 510

DB 1648 AATCAAAATGAGACTCAGAACATAGCACAGCTGCTCTTCAAAAGCACTGACAA 1707
 RESULT 15
 US-09-810-861B-3
 ; Sequence 3, Application US/09810861B
 ; Patent No. US20020162140A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Mor, Tsafir S.
 ; APPLICANT: Soreq, Hermona
 ; APPLICANT: Arntzen, Charles J.
 ; APPLICANT: Mason, Hugh S.
 ; TITLE OF INVENTION: EXPRESSION OF RECOMBINANT HUMAN ACETYLCHOLINESTERASE IN
 ; FILE REFERENCE: B71-45
 ; CURRENT APPLICATION NUMBER: US/09/810,861B
 ; CURRENT FILING DATE: 2001-03-16
 ; PRIOR APPLICATION NUMBER: 60/190,440
 ; PRIOR FILING DATE: 2000-03-17
 ; NUMBER OF SEQ ID NOS: 5
 ; SOFTWARE: PatentIn Ver. 3.1
 ; SEQ ID NO 3
 ; LENGTH: 5767
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: plasmid vector
 ; OTHER INFORMATION: pIM034.
 US-09-810-861B-3

Alignment Scores:

Pred. No.: 5,68e-43 Length: 5767
 Score: 452.00 Matches: 156
 Percent Similarity: 44.59% Conservative: 83
 Best Local Similarity: 29.10% Mismatches: 179
 Query Match: 15.78% Indels: 118
 DB: 9 Gaps: 26
 US-09-943-857-4 (1-547) x US-09-810-861B-3 (1-5767)

QY 35 AlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArgPheLysAspPro 54
 DB 1021 GCTTCTCTGGGATCCCTTTGGAGGCCACCCATGGAGCCCGCTGCTTTCTGCCACCG 1080
 QY 55 -----ValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyPro 71
 DB 1081 GAGCCCAAGCAGCCTTGGTCAGGGGTGTAGAGCTACACCTTCCAGAGTGTCTC-----1134
 QY 72 SerCysMetGln-----GlnAsnProGluGlyThrPheGluGluAsn 85
 DB 1135 ---TGCTACCATATGTGGACACCCCTATACCCAGGTTTGGAGGCCACCGAGATGTGGAAC 1191
 QY 86 LeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuPro 105
 DB 1192 CCCAACCGTGAGCTG-----1206
 QY 106 GlnSerGluAspCysLeuThrIleAsnValValArgProProGlyThrLysAlaGlyAla 125
 DB 1207 ---AGCAGGAGCTGCTGTACCTCAAC---GTGTGGACACCATACATCCCGGCTACATCC 1260
 QY 126 AsnLeuProValMetLeuTrpIlePheGlyGlyPheGluIleGlySerProThrIle 145
 DB 1261 CCCACCCCTGCTCTGTCTGGATCTATGGGGTGGCTTCTACAGTGGGCGCTCCTCTCTTG 1320
 QY 146 PheProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIleIleHis---164
 DB 1321 -----GACGTGTACGATGGCGCTTCTGTGTACAGGCC 1353
 QY 165 -----ValAlaValAsnTyrArgValAlaSerTyrGlyPheLeuAlaGly 179
 DB 1354 GAGAGGACTGTCTGTGTCCATGAACCTACCGGTGGGAGCCCTTTGGCTTCTGCGCCCTG 1413
 QY 180 AspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMet 199

Search completed: August 5, 2004, 22:32:09
Job time : 550 secs

Db	1414	CGGGGAGCCGA---GAGCCCCGGCAATGTGGTCTCTCTGGATCATGAGGCTGGCCCTG	1470
Qy	200	GlnTrpValAlaAsnIleAlaGlyPheGlyAspProSerLysValThrIlePhe	219
Db	1471	CAGTGGGTGCAGAGAACGTGCAGCCCTTCGGGGGTGACCCGACATCAGTGCAGCTGTTT	1530
Qy	220	GlyCysSerAlaGlySerMetSerValLeuCyHHisLeuIleTrpAsnAspGlyAspAsn	239
Db	1531	GGGGAGAGCGGGAGCCGCCCTCGTGGTGGCATGCACCTGCTG	1572
Qy	240	ThrTyrlsGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValPro	259
Db	1573	TCGCCGCCAGCGGGGCGCTGTTTCCACAGGGCGGTGCTGCAGAGCGGTGCC	1626
Qy	260	SerAspProValaspGlyThyTrpGly	271
Db	1627	AATGGACCTTGG---GCCACGGTGGGCATGGGAGAGGCCGCTGCAGAGCGGTGCC	1683
Qy	272	TyrAspLeuPheValSerSerAlaGlyCys	285
Db	1684	GCCACCTT-----GTGGGCTGCTCCAGCGCGCACTGGTGGGAATGACACA	1731
Qy	286	AspLysIleuAlaCysIeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsn	305
Db	1732	GAGCTGGTACCTTGCCTTCGGACACACAGCGCGAGTCTCTGGT	1782
Qy	306	ThrProGlyPheLeuAlaTySerSerLeu---AtgLeuSerTyrlsProArgProAsp	324
Db	1783	GAATGGCACGTGCTGCCTCAAGAAAGCGTCTCCGGTTCCTCTTCGTGCCTGGTAGAT	1842
Qy	325	GlyLysAsnIleThrAspAspMetTyrlsLeuValArgAspGlyLysTyrlsAlaSerVal	344
Db	1843	GGAGACTTCCTCAGTGACACCCAGAGGCCCTCATCAACGCGGGAGACTTCCACGCGCTG	1902
Qy	345	ProValIleleGlyAspGlnAsnAspGluGlyThrIlePhe	358
Db	1903	CAGGTGCTGGGTGTGGTGAAGGATGAGGGGTCTGTATTTCTGGTTTACGGGGCCCCA	1962
Qy	359	GlyLeuSerSerLeuAsn-----ValThrThrAsnAlaGlnAlaArgAlaTyrlsPheLys	376
Db	1963	GGCTTCAGCAAGACACAGAGTCTCTCATCAGCGGGCCGAGTTCCTGCGCGGGTGGCGG	2022
Qy	377	GlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaIaTyrls	394
Db	2023	GTGCGGGTTCGCCAGTAAGTCACTGGCAGCGAGGCTGTGTCTGCTGCATTACACAGAC	2082
Qy	395	-----ProGlnAsp-----IleThrGlnGlySerProPheAspThrGlyVal	408
Db	2083	TGGCTGCATCCCGAGACCCCGCAGCTCAGGAGGCCCTGAGCGATGTGGTGGCGCAC	2142
Qy	409	LeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPhe	428
Db	2143	CACATGTGCTGTGCCCC-----GTGGCCCGAGCTGGCTGGCGGAGCTGGTGGCC	2190
Qy	429	IleHisAlaArg-----ArgTyrlPheLeuAsnHisPheGlnGly	441
Db	2191	CAGGGTGGCCGGGTCTACGCTCATGCTTTGAACACCGCTGTTCCACGCTCTCTCTGGCCC	2250
Qy	442	-----GlyThrLysTyrlSerPheLeuSerLysGlnLeuSerGlyLeuProIle	457
Db	2251	CTGTGGATGGGGGTGCCCCACGGCTACGAGATCGAGTTCATCTTTGGATCCCGCTGGAC	2310
Qy	458	---MetGlyThrPheHisAlaAsnAspIleValTrpGlnAspTyrlLeuLeuGlySerGly	476
Db	2311	CCCTCTCGAAACTACACGGCACGAGGAGAAATCTTCGCCACGCGACTGATG	2361
Qy	477	SerValIleTyrlAsnAsnAlaPheIleAlaPheIleAlaThrAspLeuAspProAsnThrAla	496
Db	2362	-----CGATACTGGGCCAACTTTGCCCGCACAGGGGATCCCATGAGCCC	2406
Qy	497	-----GlyLeuLeuValAsnTrpProLysTyrlThrSerSerGln	510

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 5, 2004, 18:59:17 ; Search time 2706 Seconds
(without alignments)

6036.442 Million cell updates/sec

Title: US-09-943-857-4

Perfect score: 2864

Sequence: 1 SNNRGPAGRLGSLVPTAKLA.....DNFRTAGYDALMTNPSSFFV 547

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-O=/cgn2_1/USPTO.spool/US09943857/runat_29072004.090900.29746/app_query.fasta_1.711
-DB=EST -QFMT=fastap -SUFRTX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09943857 @CGN 1 1 2607 @runat_29072004.090900.29746 -NCFU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOC=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPO2=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOF=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hcc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	586	20.5	769	14	CF881106 trico083xn
C 2	586	20.5	824	14	CB907899 trico083xn
C 3	554.5	19.4	719	14	CF880958 trico083xc
C 4	554.5	19.4	774	14	CB907713 trico083xc
C 5	536	18.7	970	14	CF885800 trico083xn
C 6	481	16.8	1723	29	AY407738 Mus muscu
C 7	481	16.8	2173	11	AK046080 Mus muscu
C 8	481	16.8	2181	11	AK043748 Mus muscu
C 9	481	16.8	2184	11	AK042335 Mus muscu
C 10	471	16.4	3206	11	AK050337 Mus muscu
C 11	467	16.3	2026	11	AK052200 Mus muscu
C 12	461.5	16.1	2177	11	AK036259 Mus muscu
C 13	461	16.1	620	10	AW790714 D00142-R
C 14	456	15.9	604	14	CD056060 H011A1S
C 15	452	15.8	1723	29	AY407736 Homo sapi
C 16	452	15.8	1723	29	AY407737 Pan trogl
C 17	452	15.8	1745	11	BC036813 Homo sapi
C 18	452	15.8	2187	11	BC026315 Homo sapi
C 19	451	15.7	2523	14	CD014089 90134961
C 20	445.5	15.6	1800	9	AA415091 MG0026 RC
C 21	442	15.4	1958	11	AK078953 Mus muscu
C 22	442	15.4	3688	11	AK036443 Mus muscu
C 23	437.5	15.3	2527	14	CD014090 90135021
C 24	429	15.0	559	12	BM361770 A01032-R
C 25	425	14.8	1305	14	CD014075 90116512
C 26	425	14.8	1496	14	CD014071 90116512
C 27	421.5	14.7	693	14	CD054007 H002N23S
C 28	419	14.6	681	14	CD054203 H002B22S
C 29	419	14.6	2887	11	AK077248 Mus muscu
C 30	415	14.5	2661	11	AK033563 Mus muscu
C 31	409	14.3	2356	11	AK037191 Mus muscu
C 32	405	14.1	1944	11	AK078879 Mus muscu
C 33	404	14.1	1680	29	AY414461 Mus muscu
C 34	404	14.1	1918	14	CF111083 Shultzomi
C 35	404	14.1	1342	11	BC015286 Mus muscu
C 36	399	13.9	2080	11	BC001541 Homo sapi
C 37	398	13.9	4927	11	AK040349 Mus muscu
C 38	394	13.8	1194	14	CD014076 90113633
C 39	394	13.8	1364	14	CD014074 90116703
C 40	394	13.8	1585	14	CD014073 90116676
C 41	387.5	13.5	1177	14	CD0504564 CDA68-G10
C 42	387.5	13.5	1680	29	AY414459 Homo sapi
C 43	387	13.5	763	29	CG810164 F5AAQ71TF
C 44	381	13.3	670	10	BF047018 EST1115 M
C 45	379.5	13.3	939	13	BX854215 BX854215

ALIGNMENTS

RESULT 1
CF881106/c
LOCUS CF881106 769 bp mRNA linear EST 31-OCT-2003
DEFINITION trico083xn18.bl T.reesei mycelial culture, Version 6 October 2003
ACCESSION CF881106
VERSION CF881106.1 GI:381135788
KEYWORDS EST
SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM Hypocrea jecorina
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE 1 (bases 1 to 769)

AUTHORS Diener S.E., Dankmeyer, L., Dunn-Coleman, N., Houfek, T.D., Mitchell, T.K., van Solingen, P., Teunissen, P.J.M., Ward, M. and Dean, R.A.

TITLE Analysis of the protein processing and secretion pathways in a *Trichoderma reesei* EST dataset

JOURNAL COMMENT Unpublished (2003)
Contact: Ralph A. Dean
Fungal Genomics Laboratory
North Carolina State University
Campus Box 7251, Raleigh, NC 27695, USA
Tel: 919-513-0020
Fax: 919-513-0024
Email: ralph.dean@ncsu.edu
Seq primer: *IT-F1* primer:
Location/Qualifiers
source
1. 769
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/clones="trich083xn18"
/dev_stage="mycelia"
/clone_lib="T. reesei mycelial culture, Version 6 October 2003"
/note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN

Alignment Scores:
Pred. No.: 7.82e-56 Length: 769
Score: 586.00 Matches: 115
Percent Similarity: 65.90% Conservative: 28
Best Local Similarity: 53.00% Mismatches: 72
Query Match: 20.46% Indels: 2
DB: 14 Gaps: 1

US-09-943-857-4 (1-547) x CF881106 (1-769)

Qy 106 GlnSerGluAspCysLeuThrIleAsnValValArgProGlyThrIleAlaGlyAla 125
Db 652 CAGTGGGAAGACTGCTGACCATCAACATCACTGCTCCGAGGCGACGACCGGGCGAC 593

Qy 126 AsnLeuProValMetLeuTrpIlePheGlyGlyPheGluIleGly-----SerPro 143
Db 592 GCTCTGCGGGTCTCTTTTACATCTTCGAGCGGCTTCTCTCATTTGGCGGACGTGGCT 533

Qy 144 ThrIlePheProAlaGlnMetValThrIleSerValLeuMetGlyIleHisIleIle 163
Db 532 GCCCTGAACGACCCCTCCAGCTGCTCCAGCGGCGTCCGCTTCGCGAAGCCCTTCATC 473

Qy 164 HisValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGlyAspIleIle 183
Db 472 TTCGCGGGCGGTCAACTACCGCTCGCGGGTGGGGTTTCATCCCTGGGGAGAGATTCTC 413

Qy 184 AlaGluGlySerGlyAsnAlaGlyLeuIleAspGlnArgLeuGlyMetGlnTrpValAla 203
Db 412 AATGAAGGAGGCGCCACACCGCGCTTCTTGACCGAGCGCATGGGCTCGAATGGTGGCC 353

Qy 204 AspAsnIleAlaGlyPheGlyGlyAspProSerIleValThrIlePheGlyGluSerAla 223
Db 352 GATAACATTGAGCGCTTTGGCGGAACCTGCTCCGAGGTGACCATCTGGGGCGAGTCCGCT 293

Qy 224 GlySerMetSerValLeuCysHisIleIleTrpAsnAspGlyAspAsnThrTyrIleGly 243
Db 292 GGCCTCATTTCCGCTCTGGGACCGACGCTGCTGTGTATGATGGCGATGCCACATACACGAC 233

Qy 244 LysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValProSerAspProVal 263
Db 232 AAGCCTCTGTTCCGCGCGCCCATCATGAACCTCGGGTCCGTCGACGCCCGCTCGATCGGTC 173

Qy 264 AspGlyThrTyrGlyAsnGluIleTyrAspLeuPheValSerSerAlaGlyCysGlySer 283

Db 172 GATCGGAAAGGCGAAGCGGTGTATCAGCATGTCTGTCGAGAAGGCTGATGCGACGCG 113

Qy 284 AlaserAspIleuAlaCysLeuArgSerAlaserSerAspThrIleuLeuAspAlaThr 303
Db 112 AGGACGACACCTGGTGTCTCCGAACTTGACCAACGACGATTTCCGCGGACGACG 53

Qy 304 AsnAsnThrProGlyPheLeuAlaTyrSerSerIleuArgIleuSerTyrIleu 320
Db 52 AATTCGGTCCCTGGCTGATCTGATAGAGTCCCTTCTCTTTCGTACCTC 2

RESULT 2
CB907899/c
LOCUS
DEFINITION
trich083xn18 T.reesei mycelial culture, Version 3 april Hypocrea jecorina cDNA clone trich083xn18, mRNA sequence.

ACCESSION CB907899
VERSION CB907899.1 GI:30122557
KEYWORDS ESI.
SOURCE Hypocrea jecorina (anamorph: *Trichoderma reesei*)
ORGANISM Hypocrea jecorina
Bukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.

REFERENCE 1 (bases 1 to 824)
AUTHORS Foreman, P.K., Brown, D.E., Dankmeyer, L., Dean, R., Diener, S., Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J., Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C., Olivares, H.A., Teunissen, P.J., Yao, J. and Ward, M.
Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*
J Biol. Chem. 278 (34), 31988-31997 (2003)

TITLE
JOURNAL 22803314
MEDLINE 12788920
PUBMED
COMMENT Contact: Pamela K. Foreman
Genencor Intl.
925 Page Mill Road, Palo Alto, CA 94304, USA
Tel: (650) 846-7635
Fax: (650) 621-7817
Email: Pforeman@genencor.com
Seq primer: *IT-F1* primer:
Location/Qualifiers
source
1. 824
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/clones="trich083xn18"
/dev_stage="mycelia"
/clone_lib="T. reesei mycelial culture, Version 3 april"
/note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN

Alignment Scores:
Pred. No.: 8.83e-56 Length: 824
Score: 586.00 Matches: 115
Percent Similarity: 65.90% Conservative: 28
Best Local Similarity: 53.00% Mismatches: 72
Query Match: 20.46% Indels: 2
DB: 14 Gaps: 1

US-09-943-857-4 (1-547) x CB907899 (1-824)

Qy 106 GlnSerGluAspCysLeuThrIleAsnValValArgProGlyThrIleAlaGlyAla 125
Db 707 CAGTGGGAAGACTGCTGACCATCAACATCACTGCTCCGAGGCGACGACCGGGCGAC 648

Qy 126 AsnLeuProValMetLeuTrpIlePheGlyGlyPheGluIleGly-----SerPro 143
Db 647 GCTCTGCGGGTCTCTTTTACATCTTCGAGCGGCTTCTCTCATTTGGCGGACGTGGCT 588

Qy 144 ThrIlePheProAlaGlnMetValThrIleSerValLeuMetGlyIleHisIleIle 163

597 GCCCTGAACGACCCCTCCAGCTCGTCAGAGCGGCTCGCTTCCGGCAAGCCCTTCATC 528
 Db
 164 HisValAlaValAsnTyrArgValAlaSerTyrGlyPheLeuAlaGlyAspAspIleLys 183
 Qy
 527 TTCGCGCGGTCACTACCGGCTCGGCGGTGGGGTTTCATGCTGGGAGGATTC 468
 Db
 184 AlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMetGlnTyrValAla 203
 Qy
 467 AATGAAGGAGCGCAACGCGGCTTGTTCAGCAGGCGCATGGGCTCGAATGGGTGGCC 408
 Db
 204 AspAsnIleAlaGlyPheGlyAspProSerLysValThrIlePheGlyGluSerAla 223
 Qy
 407 GATAACATTCAGGCTTGGCGGAATCGTCCAGGTGACCATCTGGGGGACATCGCGT 348
 Db
 224 GlySerMetSerValLeuGlyCysHisLeuIleTyrPheAsnAspGlyAspAsnThrTyrLysGly 243
 Qy
 347 GGCTCCATTTCCGCTCTGGGACAGCTCGTGTGTATGATGGCGATGCCACATCAACGAC 288
 Db
 244 LysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValProSerAspProVal 263
 Qy
 287 AAGCCTCTGTTCGCGCGCCCATCATGAATCGGGCTTCCGTGACGCGCGTCCGATCCGGTC 228
 Db
 264 AspGlyThrTyrGlyAsnGluIleTyrAspLeuPheValSerSerAlaGlyCysGlySer 283
 Qy
 227 GACTCGAAGGAGCGGCGGTATCAGCATGTCTGAGAGGCTGGATCGACGGC 168
 Db
 284 AlaSerAspLysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThr 303
 Qy
 167 AGCGACGACACCTGGGTGTCTCCGAACTTACCAACGACGATTCGCGCGCAGCG 108
 Db
 304 AsnAsnThrProGlyPheLeuAlaTyrSerSerLeuArgLeuSerTyrLeu 320
 Qy
 107 AATTCGCTCCCTGGCTGATCTGTACGAGTCCCTTCTTTCGTACTCTC 57
 Db
 RESULT 3
 CF880958
 LOCUS
 DEFINITION
 trico83xc07.b1 T.reesei mycelial culture, Version 6 October 2003
 ACCESSION
 CF880958
 VERSION
 1 (bases 1 to 719)
 KEYWORDS
 EST.
 SOURCE
 Hypocrea jecorina (anamorph: Trichoderma reesei)
 ORGANISM
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
 REFERENCE
 1 (bases 1 to 719)
 AUTHORS
 Diener, S.E., Dankmeyer, L., Dunn-Coleman, N., Houfek, T.D.,
 Mitchell, T.K., van Solingen, P., Teunissen, P.J.M., Ward, M. and
 Dean, R.A.
 TITLE
 Analysis of the protein processing and secretion pathways in a
 Trichoderma reesei EST dataset
 JOURNAL
 Unpublished (2003)
 COMMENT
 Contact: Ralph A. Dean
 Fungal Genomics Laboratory
 North Carolina State University
 Campus Box 7251, Raleigh, NC 27695, USA
 Tel: 919-513-0020
 Fax: 919-513-0024
 Email: ralph.dean@ncsu.edu
 Seq primer: LT-Fi primer.
 FEATURES
 source
 1..719
 /organism="Hypocrea jecorina"
 /mol_type="mRNA"
 /strain="QM6a"
 /db_xref="taxon:51453"
 /clone="trico83xc07"
 /dev_stage="mycelia"
 /clone_lib="T.reesei mycelial culture, Version 6 October
 2003"
 /note="Vector: pREP3V; Site_1: Not I/Sal I; Mycelial
 culture grown from 24 hrs to 6 days with varying Carbon

and Nitrogen sources and concentrations."

ORIGIN
 Alignment Scores:
 Pred. No.: 2,79e-52 Length: 719
 Score: 554.50 Matches: 117
 Percent Similarity: 62.08% Conservative: 32
 Best Local Similarity: 48.75% Mismatches: 88
 Query Match: 19.36% Indels: 3
 DB: 14 Gaps: 1
 US-09-943-857-4 (1-547) x CF880958 (1-719)
 Qy 170 ArgValAlaSerTyrGlyPheLeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsn 189
 Db 1 CGGTTCGCGGCTTCTGCTTCTTCTGGGCGGCGGAGATCTTGGCAACGCGCAGCGCAAC 60
 Qy 190 AlaGlyLeuLysAspGlnArgLeuGlyMetGlnTyrPheGlyGluSerAlaGlyPhe 209
 Db 61 CTCGGGCTGTGACCGACGCGCATGCGGCTGCAATGGGTGGCGACATCGCGGCTTT 120
 Qy 210 GlyGlyAspProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeu 229
 Db 121 GCGCGGCGACCGGACCAAGGTGACCATCTGGGGGAGTTCGCGCGGCGCATGTCCGTCTAC 180
 Qy 230 CysHisLeuIleTyrPheAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAla 249
 Db 181 GACCAGATGTGTGTACGCGGCAACGCGCATACAGCGCAAGCCCTGTTCGCGCGC 240
 Qy 250 GlyIleMetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsn 269
 Db 241 GCATCATGAATCTTGCGACGCGGCTGCGCGGCGGCGGCTGACTCGCGCAGGCGCAG 300
 Qy 270 GluIleTyrAspLeuPheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAla 289
 Db 301 GCCATCTACGACCAACGCTCGTGAAAAAGGCGGCTGCGCGGCTCCGCGCGCATGCTGGCG 360
 Qy 290 CysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPhe 309
 Db 361 TGCTTCGCGCAGGTGCGCTTACAGCAAGTTCCTCGCGCGGTCAACGTCGCGCGGCGCATC 420
 Qy 310 LeuAlaTyrSerSerLeu-ArgLeuSerTyrLeuProArgProAspGlyLysAsnIleT 329
 Db 421 CTGTCTACAGCTCTCTGGGGGCTGTCTACTGCGCGCGCGCGGCGGCTGCTCTCTCC 480
 Qy 329 hrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleIleG 349
 Db 481 CGGACAGCGCGCGGAAAGCGCTCTCTGACGAGGCGAGGTACCAAGCGCGTCCCATGATGCGG 540
 Qy 349 LysAspGlnAsnAspGluGlyThrIlePheGlyLysSerSerLeuAsnValThrThrAsnA 369
 Db 541 GCGACCGAGGACGAGGCGACGCTTCTTCATCTTCCAAAGCGCGCTCAACAGCAGCGC 600
 Qy 369 laGlnAlaArgAlaTyrPheLysGln--SerPheIleHisAlaSerAspAlaGluIleA 388
 Db 601 CGAGCTCTGTGACCTGGCTCTCGGAGTACATGTTTCCAGGCGGCGCGCGAGCTGA 660
 Qy 388 spThrLeuMetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThr 406
 Db 661 CGCAGCTCGTCAACACCTACGACACGCGNCACTCCAGCGCGCGCGCCCTTTCGACG 716
 RESULT 4
 CF880958
 LOCUS
 DEFINITION
 trico83xc07 T.reesei mycelial culture, Version 3 april Hypocrea
 jecorina cDNA clone trico83xc07, mRNA sequence.
 ACCESSION
 CF880958
 VERSION
 1 (bases 1 to 719)
 KEYWORDS
 EST.
 SOURCE
 Hypocrea jecorina (anamorph: Trichoderma reesei)
 ORGANISM
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.

```

REFERENCE
AUTHORS
  1 (bases 1 to 774)
Foreman,P.K., Brown,D.E., Dankmeyer,L., Dean,R., Diener,S.,
Dunn-Coleman,N.S., Goedegebuur,F., Houfek,T.D., England,G.J.,
Kelley,A.S., Meerman,H.J., Mitchell,T., Mitchinson,C.,
Olivares,H.A., Teunissen,P.J., Yao,J. and Ward,M.
Transcriptional regulation of biomass-degrading enzymes in the
filamentous fungus Trichoderma reesei
J. Biol. Chem. 278 (34), 31988-31997 (2003)
JOURNAL
MEDLINE
PUBMED
  12788920
COMMENT
  Contact: Pamela K. Foreman
  Genencor Intl.
  925 Page Mill Road, Palo Alto, CA 94304, USA
  Tel: (650) 846-7635
  Fax: (650) 621-7817
  Email: Pforeman@genencor.com
  Seq primer: LT-F1 primer.
  Location/Qualifiers
    1..774
      /organism="Hypocrea jecorina"
      /mol_type="mRNA"
      /strain="QM6a"
      /db_xref="taxon:51453"
      /clone="trich083xc07"
      /dev_stage="mycelia"
      /clone_lib="T.reesei mycelial culture, Version 3 april"
      /note="Vector: pREP3Y, Site 1: Not I/Sal I; Mycelial
      culture grown from 24 hrs to 6 days with varying Carbon
      and Nitrogen sources and concentrations."
ORIGIN
source
Alignment Scores:
Pred. No.: 3.17e-52 Length: 774
Score: 554.50 Matches: 117
Percent Similarity: 62.08% Conservative: 32
Best Local Similarity: 48.75% Mismatches: 88
Query Match: 19.36% Indels: 3
DB: 14 Gaps: 1
US-09-943-857-4 (1-547) x CB907713 (1-774)
QY 170 ArgValAlaSerTirGlyPheLeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsn 189
Db 56 CGCGTCGCGGCTTCTGCTTCTGGGGCGCGGAGATCTCGCAACGGCAGCGCAAC 115
QY 190 AlaGlyLeuLysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPhe 209
Db 116 CTCGGGCTGCTGGACCGCGCATGGGCTTGCAGTGGGTGGCGGACAAACATCGCGCCTTT 175
QY 210 GlyGlyAspProSerIysValThrIlePheGlyGluSerAlaGlySerMetSerValLeu 229
Db 176 GCGCGCGACCCCGGACAGGTGACCACTCTGGGGCGAGTCCGCGGCGCCATGTCGCTCTAC 235
QY 230 CysHisLeuIleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAla 249
Db 236 GACCAGATGGTGTACGGGGCGACCGCACGATACACGGCAAGCCCTGTCGCGGC 295
QY 250 GlyIleMetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsn 269
Db 296 GCATCATGAACCTCTGCACCGCGGTGCGCGCGAGCGGTGCACTCGCCCAAGCGCGAG 355
QY 270 GluIleTyrAspLeuPheValSerAlaGlyCysGlySerAlaSerAspLysLeuAla 289
Db 356 GCATCTACGACACGTCGTGAAAAAGCGCGGTGAGCGCTCCGCGACTCGCTGGCG 415
QY 290 CysLeuArgSerAlaSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPhe 309
Db 416 TGCCTCGCCAGGTGCCTACAGCAAGTCTCTCGCGCGCGTCAACGTCGCGCGCGGCATC 475
QY 310 LeuAlaTyrSerSerLeu--ArgLeuSerTyrLeuProArgProAspGlyLysAsnIle 329
Db 476 CTGTCTTACAGCTCCCTGGGGGGTGTCTTACTTCTCGCGCGCGCGCGCGCTGTCTCTCC 535
QY 329 hrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleLeu 349
Db 536 CGGACACGCCCGGAAGCGCTCTCGAGCAGGCGAGGTACACGCGGTGCCCATGTTCCGC 595
QY 349 LysAspGlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnA 369
Db 596 GCACACGAGGAGGAGGCGACGCTCTCTCCATCTTCCAAAGCGAGCTCAACAGCAGCG 655
QY 369 laGlnAlaArgAlaTyrPheLysGln---SerPheIleHisAlaSerAspAlaGluLeuA 388
Db 656 CGAGCCTCGTGACCTGGCTCTCGAGTACATGTTTCCAGGGGACGCGCGCGAGCTGA 715
QY 388 spThrLeuMetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThr 406
Db 716 CGCAGCTCGTCAACACCTACGACACGCGNCATCTCCAGCGGAGCCCTTTTCGCACG 771
RESULT 5
CF885800/c 970 bp mRNA linear EST 31-OCT-2003
LOCUS trich083xn16.b11 T.reesei mycelial culture, Version 6 October 2003
DEFINITION Hypocrea jecorina cDNA clone trich083xn16, mRNA sequence.
ACCESSION CF885800
VERSION CF885800.1 GI:38140482
KEYWORDS EST.
SOURCE Trichoderma reesei
ORGANISM Hypocrea jecorina (anamorph: Trichoderma reesei)
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE
  1 (bases 1 to 970)
Diener,S.E., Dankmeyer,L., Dunn-Coleman,N., Houfek,T.D.,
Mitchell,T.K., van Solingen,P., Teunissen,P.J.M., Ward,M. and
Dean,R.A.
Analysis of the protein processing and secretion pathways in a
Trichoderma reesei EST dataset
Unpublished (2003)
COMMENT
  Contact: Ralph A. Dean
  Fungal Genomics Laboratory
  North Carolina State University
  Campus Box 7251, Raleigh, NC 27695, USA
  Tel: 919-513-0020
  Fax: 919-513-0024
  Email: ralph.dean@ncsu.edu
  Seq primer: LT-F1 primer.
  Location/Qualifiers
    1..970
      /organism="Hypocrea jecorina"
      /mol_type="mRNA"
      /strain="QM6a"
      /db_xref="taxon:51453"
      /clone="trich083xn16"
      /dev_stage="mycelia"
      /clone_lib="T.reesei mycelial culture, Version 6 October
      2003"
      /note="Vector: pREP3Y, Site 1: Not I/Sal I; Mycelial
      culture grown from 24 hrs to 6 days with varying Carbon
      and Nitrogen sources and concentrations."
ORIGIN
Alignment Scores:
Pred. No.: 6.17e-50 Length: 970
Score: 536.00 Matches: 120
Percent Similarity: 55.60% Conservative: 34
Best Local Similarity: 43.32% Mismatches: 108
Query Match: 18.72% Indels: 16
DB: 14 Gaps: 3
US-09-943-857-4 (1-547) x CF885800 (1-970)
QY 48 AsnLeuArgPheLysAspProValProTyrSerGlySerLeuAsnGlyGlnLysPheThr 67
Db 798 AACCTCGGTGGACACCAACCAATGTGCTAGTGTCCATT----- 757
QY 68 SerTyrGlyProSerCysMetGlnGlnAsnProGluGlyThrPheGluGlnLeuGly 87

```

```

Db 756 -----GCATTCGCGCGCCCAATCATCAGTAGCACAAATTCGTCC 712
Qy 88 LysThrAlaLeuAspLeuValMetGlnSer-----LysValPheGlnAlaValLeuPro 105
Db 711 CCGCCACGCTGGCAGAGTGTCAATTAACAGGCGCTGCTTTAGCGGAGCGAGCCA 552
Qy 106 GlnSerGluAspCysLeuThrIleAsnValValArgProGlyThrLysAlaGlyAla 125
Db 651 GT-AGTAAATGCAATACATACAGCTTCGCTAGTCCAAAGGCGCGCCACGCGGCGAC 593
Qy 126 AsnLeuProValMetLeuTrpIlePheGlyGlyPheGluIleGly-----SerPro 143
Db 592 GCTTCGCGGTCTTTTACACTCTCGAGCGCGCTTCGCTATGCGGAGCGTCCGTT 533
Qy 144 ThrIlePheProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIleIle 163
Db 532 GCCTAAACGACCCCTCCAGCTCGTCAGACGCGCGTTCGCTTCGCGACGCTTTCATC 473
Qy 164 HisValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGlyAspIleLys 183
Db 472 TTCGCGCGGTCACTACCGCGTGGGCGGTGGGTTTCATGCTGGGAGGAGATTC 413
Qy 184 AlaGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMetGlnTrpValAla 203
Db 412 AATGAGGAGCGCCAAACGCGCTTCGTTGACGAGCGCATGGGCGCTCGAATGGTGGCC 353
Qy 204 AspAsnIleAlaGlyPheGlyGlyAspProSerLysValThrIlePheGlyGluSerAla 223
Db 352 GATAACATTCGAGCGCTTCGCGGGAACCTCGTCGAGGTGACCATCTGGGCGCATCCGCT 293
Qy 224 GlySerMetSerValLeuCysHisLeuIleTrpAsnAspGlyAspAsnThrTyrLysGly 243
Db 292 GGCTCCATTTCCGCTCTGGGACCACTGCTGCTGATGCGCATGCCATACATCAACGAC 233
Qy 244 LysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValProSerAspProVal 263
Db 232 AAGCTCTGTTCGCGCGCCATCATGAACCTCGGCTCGCTGACGCGCCGTCGATCCGCTC 173
Qy 264 AspGlyThrTyrGlyAsnGluIleTyrAspLeuPheValSerSerAlaGlyCysGlySer 293
Db 172 GACTCGAAAGGCAAGCGGTGATCAGCATGTCTCGAAGAGCTGTGATCGGAGCGC 113
Qy 284 AlaSerAspLysLeuAlaCysLeuArgSerAlaSerAspThrLeuLeuAspAlaThr 303
Db 112 AGCGAGCACCCCTGGGTGTCTCCGGAACCTTGACCAACGACGATTCGCGCGAGCG 53
Qy 304 AsnAsnThrProGlyPheLeuAlaTyrSerSerLeuArgLeuSerTyrLeu 320
Db 52 AATTCGCTCCCTGGCTGATCTCGTACGAGTCCCTTGTCTTTCGTACCTC 2

RESULT 6
LOCUS AY407738
DEFINITION Mus musculus ACHE gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY407738.1 GI:39763709
VERSION 1 (bases 1 to 1723)
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1723)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B.,
Adams,M.D., and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1723)

```

```

AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
location/Qualifiers
1..1723
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..>1723
/gene="ACHE"
/locus_tag="HMC2990"
gene
ORIGIN
Alignment Scores:
Pred. No.: 3,32e-43 Length: 1723
Score: 481.00 Matches: 162
Percent Similarity: 44.69% Conservative: 78
Best Local Similarity: 30.17% Mismatches: 177
Query Match: 16.79% Indels: 120
DB: 29 Gaps: 26
US-09-943-857-4 (1-547) x AY407738 (1-1723)
Qy 35 AlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArgPheLysAspPro 54
Db 184 GCITTTCTGGGATCCCTTTGAGAGCCACCTGTGGCTCAGTAGATTATGCCACCA 243
Qy 55 Val-----ProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyPro 71
Db 244 GAGCCCAAGCGCGCTGGTCAGGAGTGTGGATGCT-----ACCACCTTCCAAAT 294
Qy 72 SerCysMetGln-----GlnAsnProGluGlyThrPheGluGluAsn 85
Db 295 GTCTGTACAGTACGTGGACCCCTGTACCTGGGTTTGGGGTACTGAGATGTGGAAC 354
Qy 86 LeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuPro 105
Db 355 CCCAACCGAGAGTTG----- 369
Qy 106 GlnSerGluAspCysLeuThrIleAsnValValArgProGlyThrLysAlaGlyAla 125
Db 370 ---AGTGAAGACTGCTGTATCTTAATGTGTGGACACCATACCCAGACCTGCTTCTCC 426
Qy 126 AsnLeuProValMetLeuTrpIlePheGlyGlyPheGluIleGlySerProThrIle 145
Db 427 ACA---CCTGTCTCATCTGGATCTATGGGGTGGTTCACAGCGGAGCGGCTCTT 483
Qy 146 PheProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIle----- 162
Db 484 -----GATGTGTATGACGCGCGTTCCTGCCCGAGTT 516
Qy 163 -----IleHisValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGly 179
Db 517 GAGGGAGCTGTGTGGTATCTATGAACCTACCGAGTGGGAACCTTTGGCTTCTTGGCCCTA 576
Qy 180 AspAspIleLysAlaGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMet 199
Db 577 CCAGGAACGAGA---GAAGCCCTGGCAATGTAGTCTGTGGATCAACGGCTTCCTTG 633
Qy 200 GlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysValThrIlePhe 219
Db 634 CAATGGGTGCAGAAAATATTGCAGCCTTTGGGCGGACCCGATGTCAGTACTGCTGTT 693
Qy 220 GlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAspGlyAspAsn 239
Db 694 GGGGAGAGTGGCGGTGGAGCTCCGCTGGCGATGCACATAGT----- 735

```



```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:B230340L13"
/db_xref="MGI:2411467"
/db_xref="taxon:10090"
/clone="B230340L13"
/sex="male"
/tissue_type="corpora quadrigemina"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
108_1952
/note="unnamed protein product; acetylcholinesterase
(MGI:87876, GB|NM_009599, evidence: BLASTN, 99%,
match=1845)
putative"
/codon_start=1
/protein_id="BAC32595.1"
/db_xref="GI:26337819"
/translation="MRPPWYPLHTPSLAPFLLLSLGGARAEGRDPQLVVRV
GGQLRGLRKPAGPVSAFLGIPFAEPVPVGRFPPEPRKPMWSGVLDATTFONVQYQ
YVDTLPGFEGTEGMMNPRLSDCLVNWMTYPRPAPTPVLIIWVGGEYSRAS
LDVYDRFLAQVQGVAVLVSVMYRVGTFGLAPGSREAPCNVGLDORLALOWQENI
AAFQGDPMSTVTLFEGSAGASVGMHILSLPSRSLFRAVLQSTPNQWATVSAGAR
RRATLLARLVGPPGAGNDIELIACLRTPAQDLVDHEWHVLPQESIFRFSFVFW
DGFUSDTEALINTGDFODLQVLGVVKDEGSYFLVYGVPGFSKONESLISRAQLA
GVRIQVQASDLAAEAVLHLYTDLHPEDPETHLRDAMSAVGVGHNVCPVAQLAGRLA
ACQARVYAIIFHRASTLTWPLMMGVPHGYEIEFGLPLDPSLNTTTERIFAPQRLM
KYWTNPARTGDPDRDSKPPWPVYTTAAQOVSLNKLKPLEVRGLRAQTCAPNRF
LPKLLSATDTLDEAEQKAEFRHWSSVYVHMKNQFDHYSKQERCSDL"
2158_2163
/note="putative"

```

polyA_signal

polyA_site

ORIGIN

Alignment Scores:

```

Pred. No.: 5e-43 Length: 2173
Score: 481.00 Matches: 162
Percent Similarity: 44.69% Conservative: 78
Best Local Similarity: 30.17% Mismatches: 177
Query Match: 16.79% Indels: 120
DB: 11 Gaps: 26

```

US-09-943-857-4 (1-547) x AK046080 (1-2173)

```

QY 35 AlapheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArgPheLysAspPro 54
DB 291 GCTTTTCTGGGATATCCCTTTCAGAGCCACCTGTGGGCTCAGTAGATTATGCCACCA 350
QY 55 Val-----ProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyPro 71
DB 351 GAGCCCAAGCGGCCCTGGTTCAGAGTGTGGATGCT-----ACCACCTTCCAAAT 401
QY 72 SerCysMetGln-----GlnAsnProGluGlyThrPheGluGluAsn 35
DB 402 GTCTCTACCATGACCTGGACACCCCTGTACCTGGGTTTGAGGGTACTGAGATGTGGAAC 461
QY 86 LeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuPro 105
DB 462 CCCAACCGAGATTG----- 476
QY 106 GlnSerGluAspCysLeuThrIleAsnValValArgProProGlyThrLysAlaGlyAla 125
DB 477 ---AGTGAAGACTGCTGTATCTTAATGTGTGGACACCATACCCAGACCTGCTCTCC 533
QY 126 AsnLeuProValMetLeuThrIlePheGlyGlyGlyPheGluIleGlySerProThrIle 145
DB 534 ACA---CTGTCTCTCATCTGATCTATGCGGGTGTGTTCTACAGCGGAGCGCCCTCTTG 590
QY 146 PheProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIle----- 162
DB 591 -----GATGTGTATGACGCGCCGTTCTCTGGCCCGAGGTT 523

```

```

QY 163 -----IleHisValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGly 179
DB 624 GAGGAGCTGTGTGGTATCTATGACTACCGATGGAAACCTTTGGCTTCTTGGCCCTA 683
QY 180 AspAspileLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMet 199
DB 684 CCAGGAAGCAGA---GAAGCCCTGGCAATGTAGGTCTGTGATCAACGGCTTGCCTTG 740
QY 200 GlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysValThrIlePhe 219
DB 741 CAATGGTGCAGAAAATATTGCAGCCTTTGGGGCGACCCGATGTCACTGACTCTCTTT 800
QY 220 GlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAspGlyAspAsn 239
DB 801 GGGGAGAGTGGCGGTGCAGCTCCGTGGGCATGCACATCTG----- 842
QY 240 ThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValPro 259
DB 843 TCCTTCGCCAGCAGGAGGCTCTTCCACAGGGCTGTCTCCAGAGTGGC-----ACACCC 896
QY 260 SerAspPro-----ValAspGlyThrTyrGlyAsnGluIleTyrAspLeuPheVal 276
DB 897 AATGGGCCCTGGGCCACTGTGAGTGTGGAGAGCGCCAGCGCCACACTGCTGGCC 956
QY 277 SerSerAlaGlyCys-----GlySerAlaSerAspLysLeuAlaCys 290
DB 957 CGCCTTGTGGCTGTCCCCAGCGTGGCTGTGGCAATGACACCGAGTGTAGGCTGC 1016
QY 291 LeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeu 310
DB 1017 TTGAGGACAAAGCCCGCTCAGGACCTGGTGGACCAAGTGGCACGCTC-----CTG 1067
QY 311 AlaTyrSerSerLeu---ArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThr 329
DB 1068 CTTCAAGAAAGATCTCTCGGATTTCTTCGTGCTGTGTAGAGCGGACTTCCTCAGT 1127
QY 330 AspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleIleGly 349
DB 1128 GACACACCGGAGCTCTCATCAATCTGGAGATTTTCAAGACCTCAGGTGTGTGGGT 1187
QY 350 AspGlnAsnAspGlyThrIlePhe-----GlyLeuSerSerLeu 363
DB 1188 GTGGTGAAGGACGAGGGCTCCTACTTTCTGGTTTACGGGGTCCAGGCTTCAGCAAGAC 1247
QY 364 Asn-----ValThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPheIleHis 381
DB 1248 AATGAATCTCTCATCAGCGCGGCCAGTTCTCTGGCTGGGGTTCGGATCGGTGTACCCCAA 1307
QY 382 AlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyr-----ProGln 396
DB 1308 GCAAGTGTACCTGGCAGCGAGCTGTGTCTGTGCTATACAGACTGGCTGGCACCCTGAG 1367
QY 397 AspIleThrGlnGlySerProPheAspThrGlyValLeu-----AsnAlaLeuThr 413
DB 1368 GACCTCTACCTCAGAGATGCCATGATGTCAGTGTAGGCGACCAACAGTGTGTGTC 1427
QY 414 ProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHisAlaArgArg 433
DB 1428 CCT-----GTGGCCAGCTGGCTGGCGGACTGGCT----- 1457
QY 434 TyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSerLysGlnLeuSer 453
DB 1458 -----GCCCAAGGGCGCGGGTCTATGCTCTACATCTTTGACACCGTGC 1502
QY 454 GlyLeu-----ProIle---MetGlyThrPheHisAlaAsnAspIleValTrpGln 469
DB 1503 TCCACACTGACTTGGCCCTCTGGATGGGGTGGCCCAATGGCTATGAATC----- 1553
QY 470 AspTyrLeuLeuGly-----SerGlySerValIleTyrAsnAlaPheIleAla 486
DB 1554 GAGTTCATCTTTGGGCTCCCTCGATCCCTCGTGAATACACACCGAGGAGGATC 1613

```

QY 487 PheAlaThrAspLeu-----AapProAsnThr 495
 Db 1614 TTTCCTCAGGACTTATGAATACTGACCAATTTGCCCGCAGGCGGACCCATGAC 1673
 QY 496 Ala-----GlyLeuLeuValAlaThrProGlyeTyThrSerSerSerGln 510
 Db 1674 CCTCGAGACTCAAAATCTCCACAGTGCCACCGTACACCACTGCGGCGAG 1724

RESULT 8
 LOCUS AK043748
 DEFINITION Mus musculus 10 days neonate cortex cDNA, RIKEN full-length enriched library, clone:A830027P05 product:acetylcholinesterase, full insert sequence.

ACCESSION AK043748.1 GI:26335880
 VERSION AK043748.1
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1
 Carninci,P. and Hayashizaki,Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636

REFERENCE 2
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159

REFERENCE 3
 Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A., Yamamoto,K., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,S., Watahiki,M., Yoneda,Y., Iihikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
 TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913
 PUBMED 11076861

REFERENCE 4
 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409, 685-690 (2001)
 MEDLINE 11076861

REFERENCE 5
 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
 TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 JOURNAL Nature 420, 563-573 (2002)
 MEDLINE 120530913
 PUBMED 120530913

REFERENCE 6
 Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Inotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kurihara,C., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koyama,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,K., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
 TITLE Direct Submission

JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
 COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Please visit our web site for further details.
 URL:http://genome.gsc.riken.go.jp/
 URL:http://fantom.gsc.riken.go.jp/.

FEATURES
 Location/Qualifiers
 1..2181
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM_DB:A830027P05"
 /db_xref="MG1:2408986"
 /db_xref="taxon:10090"
 /clone="A830027P05"
 /tissue_type="cortex"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="10 days neonate"
 111..1955
 /note="unnamed protein product; acetylcholinesterase (MG1:87876, GB|NM_009599, evidence: BLASTN, 99%, match=1845)
 putative"
 /codon_start=1
 /protein_id="BAC31641.1"
 /db_xref="GI:26335881"
 /translation="MRPPVPLHTPPLSLAFLLLSLGGARAEGRDQLLVRVR GGQLRLKAPGPGVSAFLGIPAEPPVGSRRMPPEKPPKPSGVLDTFTQVVCY LDVLYPGFEGTEWNNRNRLSEDCLYNVNRPVPASPTPVLITVYGGFYSGAAS AAFGGDPMVTLFEGSAGASVGHILSLPSRLFRAVLOSGLTGNPFWATVSAGEAR RRAATLRLVCGPEPGAGNDTELIACLTRPAQDLVDHEHVLVLPQESIFRFPVVR DGDFLTPALINTGDFQLQVLMVDPEDBSGYLVYGVGFKDNESLTSRAQFLA GVRIGVPSQASDAEAVALVHTDMLHPEDHSDLRDAMSADVGDHNVVCPVLAQRLA AQARVAYIFEPHASTLTWPLWGVPHGYEIEFGLPLDPSLNTYTERIFACRLM KYTNFARTGDPNDRPSKSPQWPEPYTTAAQVYSLNKLPLEVRGLRAQTCAPWNR LKLLSATDLDLAEEROMKAEFFHRWSSTVHWKQFDRHYSKQERCSDL"
 2161..2166
 /note="putative"
 2181
 /note="putative"

polyA_signal
 polyA_site

ORIGIN
 Alignment Scores:
 Pred. No.: 5,03e-43 Length: 2181
 Score: 481.00 Matches: 162
 Percent Similarity: 44.69% Conservative: 78
 Best Local Similarity: 30.17% Mismatches: 177
 Query Match: 16.79% Indels: 120
 DB: 11 Gaps: 26

US-09-943-857-4 (1-547) x AK043748 (1-2181)
 QY 35 AlaPheLeuGlyLeuProPheAlaGluProValGlyAsnLeuArgPheLysAspPro 54
 Db 294 GCTTTTCTGGGATCCCTTTTCAGAGCCACCTGCGGCTCAGGTAGATTATGCCACCA 353
 QY 55 Val-----ProTySerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyPro 71
 Db 354 GAGCCCAACGGCCCTTGTCAGAGTGTGGATGCT-----ACCACCTTCCAAAAT 404
 QY 72 SerCysMetGln-----GlnAsnProGluGlyThrPheGluGluAsn 85
 Db 405 GTCTGCTACCACTACGTGGACACCCCTGTACCTGGGTTTGGGGTACTGAGATGTGAAC 464

QY	86	LeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuPro	101
DB	465	CCCAACCGAGATTG-----	479
QY	106	GlnSerGlyAspCysLeuThrIleAsnValValArgProProGlyThrLysAlaGlyAla	125
DB	480	---AGTGAAGACTGCTGTATCTTAATGTGTGGACACCATACCCAGACCTGCTTCTCC	536
QY	126	AsnLeuProValMetLeuThrIlePheGlyGlyPheGluIleGlySerProThrIle	145
DB	537	ACA---CTGTCTCATCTGATCATATGGGGTGGTTTTCACAGCGAGCGGCTCTTGT	593
QY	146	PheProProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIle	162
DB	594	-----CATGTGTATGACGGCGTTTCTTGCCCCAGGTT	626
QY	163	-----IleHisValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGly	179
DB	627	GAGGAGCTGTGTGGTATCTATGAATACCCAGGTGGGAACCTTTGGCTTCTTGCCCTA	686
QY	180	AspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMet	199
DB	687	CCAGAGACAGA---GAAGCCCTGGCAATAGTCTCTGGATCAACGGCTTGGCTT	743
QY	200	GlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysValThrIlePhe	219
DB	744	CAATGGGTGCAAGAAATATTGCAGCCTTTGGGGCGGACCCGATGTCTGACTCTGTT	803
QY	220	GlyGlnSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAspGlyAspAsn	239
DB	804	GGGAGAGTGGGGTGCGAGCTCCGTGGGCATGCACATACG-----	845
QY	240	ThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValPro	259
DB	846	TCCCTGCCACGAGGAGCTCTCCACAGGGCTGTCTCCAGATGGC-----ACACCC	899
QY	260	SerAspPro-----ValAspGlyThrTyrGlyAsnGluIleTyrAspLeuPheVal	276
DB	900	AATGGGCCCTGGGCCACTGTGAGTCTGGAGAGGCCAGGCGGCGCACACTGCTGGCC	959
QY	277	SerSerAlaGlyCys-----GlySerAlaSerAspLysLeuAlaCys	290
DB	960	CGCCTTGTGGCTGTCTCCCCAGGTGGCGCTGTGGCAATGACCCAGCTGATAGCTGC	1019
QY	291	LeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeu	310
DB	1020	TTGAGGACAGGCCCGCTCAGGACCTCGTGGACACGAGTGGCACGTC-----CTG	1070
QY	311	AlaTyrSerSerLeu---ArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThr	329
DB	1071	CCTCAAGAAAGTATCTTCCGATTTCTCGTGGCTGTGTAGACGGGGACTTCTCAGT	1130
QY	330	AspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleIleGly	349
DB	1131	GACACACCGAGGCTCTCATCAATACTGGAGATTTTCAAGACCTGCAGGTGCTGGTGGT	1190
QY	350	AspGlnAsnAspGluGlyThrIlePhe-----GlyLeuSerSerLeu	363
DB	1191	GTGGTAGAGACGAGGCTCTTACTTCTGGTTTACGGGTCCTCAGGCTTCACGAAGAC	1250
QY	364	Asn-----ValThrThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPheIleHis	381
DB	1251	AATGAATCTCTCATCAGCCGGGCCAGTTCTTGCTGGGTGGGATCGGTGTACCCCAA	1310
QY	382	AlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyr-----ProGln	396
DB	1311	GCAAGTGACCTGCAGCGGAGGTGTGGTCTCTGCAITTACACAGACTGGCTGCACCTGAG	1370
QY	397	AspIleThrGlnGlySerProPheAspThrGlyValLeu-----AsnAlaLeuThr	413
DB	1371	GACCTTACTCACTGAGAGATGCCATGTGAGTGTGATGGCGGACCAACATGTTGTGTC	1430

QY	414	ProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHisAlaArgArg	433
DB	1431	CCT-----GTGGCCAGTGCCTGGCGCAGTGCT-----	1460
QY	434	TyrPheLeuAenHisPheGInglyGlyThrLysTyrsrPheLeuSerLysGlnLeuSer	453
DB	1461	-----GCCAAGGGGCCGGGTCTATGCCTACACTTTTGAAACCACCGTCGC	1505
QY	454	GlyLeu-----Profile--MetGlyThrPheHisAlaAsnAlaPheIleValTrpGln	469
DB	1506	TCCACACTGACTTGGCCCTCTGGATGGGGTGCCCATGCTATGAATC-----	1556
QY	470	AspTyrLeuLeuGly-----SerGlySerValIleTyrAsnAsnAlaPheIleAla	486
DB	1557	GAGTTTCATTCTTGGGCTCCCCTGGATCCCTCGCTGAACCTACACCGAGGAGGATC	1616
QY	487	PheAlaThrAspLeu-----AspProAnThr	495
DB	1617	TTTTCTCAGCGACTTATGAATACTGACCAATTTTCCCGCACAGGGACCCCAATGAC	1676
QY	496	Ala-----GlyLeuLeuValAsnTyrPrroLysTyrThrSerSerSerGln	510
DB	1677	CCTCGAGACTCCAATCTCCACAGTGGCCACCGTACACCACTCCCGCGCAG	1727
RESULT	9		
LOCUS	AK042335		
DEFINITION	Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched library, clone:A630083G07 product:acetylcholinesterase, full insert sequence.	2184 bp mRNA linear HTC 19-SEP-2003	
ACCESSION	AK042335		
VERSION	AK042335.1 GI:26333054		
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	1	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)	
AUTHORS	2	20499374	
TITLE	3	11042159	
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE	4	11076861	
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			

of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 6 (bases 1 to 2184)

JOURNAL
 REFERENCE
 AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M.,
 Hayaehida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
 Takeda, Y., Tanaka, I., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.
 Direct Submission

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Sushiro-cho, Teurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.

Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
 Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
 Trust/MRC building Addenbrookes Hospital Cambridge) whose
 assistance we gratefully acknowledge.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/
 URL: http://fantom.gsc.riken.go.jp/
 Location/Qualifiers

1. 2184
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM DB:A630083G07"
 /db_xref="MGI:2407367"
 /db_xref="taxon:10090"
 /clone="A630083G07"
 /tissue_type="thymus"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="3 days neonate"
 115..1959
 /note="unnamed protein product; acetylcholinesterase
 (MGI:MGI:87876, GB|NM_009599, evidence: BLASTN, 99%,
 match=1845)
 putative"
 /codon_start=1
 /protein_id="BAC31228.1"
 /db_xref="GI:26335055"
 /translation="MRPPVPLHTPSLAPFLILLILGGGARGRBDPOLLVRVR
 GGLRLKARFGVSAFLGIPPAEPVGSRRPPEPRKPMGSLVDAITTFONVCYO
 YVDLYPFGTEGMNPNRSEDCILNVWTPYPRPASPTVLIIWYGGYSGAAS
 LVYDGRFLAQVEGAVLVSMYRVGTFFGLPLPGSREAPNVGLLDQLALQWGENI
 AAFGGPMSTVLPFGESAGASVGMHILSLPSRLPHRAVLQSGTPNGPWATVACEAR
 REATLLRLVGGCPGGAGNDTELLACLTRPAQDLVHREHVLVQESIFRFSFVFW
 DGDPLSDTEALINTGDFODLVLVGVVDEGYSFLVYGVPGFSGKONESLISRAQFLA
 CVRIQVQASDLAAEAVALHYTMDLHPEDTHLRDAMSAVVDGHNVCVPAOLAGRLA
 AQARVYATIFEHRASTLTWPLMMGVPHGYEIEIFGLPLDPLSNLTTERIFAPRLM
 KYWTNFARTGDPNDPDSKSPQMPFITTAQQQVLSNLKPLSVLRGLRQTCAPNRF
 LPKLASDTLDEAEQKAEFHRWSSYNVHWKMQFDHYSKQERCSDL"
 2165..2170
 /note="putative"
 2184
 /note="putative"

polya_signal
 polyA_site
 ORIGIN
 Alignment Scores:

Pred. No.:	5,04e-43	Length:	2184
Score:	481.00	Matches:	162
Percent Similarity:	44.69%	Conservative:	78
Best Local Similarity:	30.17%	Mismatches:	177
Query Match:	16.79%	Indels:	120
DB:	11	Gaps:	26
US-09-943-857-4 (1-547) x AK042335 (1-2184)			
QY	35	AlaPheLeuGlyIleProPheAlaGluProValGlyAsnLeuArgPheLeuAspPro	54
Db	298	GCITTTCTGGGCATCCCTTTTGAGAGCCACCTGTGGCTCACGTAGATTATGCCACCA	357
QY	55	Val-----ProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyPro	71
Db	358	GAGCCCAAGCGCCCTGGTCAGGAGTGTGGATGCT-----ACCACCTTCCAAAT	408
QY	72	SerCysMetGln-----GlnAsnProGluGlyThrPheGluGluAsn	85
Db	409	GTCTGTACCACTACGTGGACACCTGTACCTGGGTTTGGGGTACTGAGATGTGGAC	468
QY	86	LeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuPro	105
Db	469	CCCACCGAGAGTTG-----	483
QY	106	GlnSerGluAspCysLeuThrIleAsnValValArgProGlyThrLysAlaGlyAla	125
Db	484	---AGTGAAGACTGCCTGTATCTTAATGTGTGGACACCATACCCAGACCTGCTTCTCC	540
QY	126	AsnLeuProValMetLeuTyrPheGlyGlyPheGluIleGlySerProThrIle	145
Db	541	ACA---CCTGTCTCATCTGGATCTATGGGGTGTGTTCTACAGCGAGCGCCCTCTGT	597
QY	146	PheProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIle-----	162
Db	598	-----GATGTGTATGACGCGCTTCTGTGCGCCAGGTT	630
QY	163	-----IleHisValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGly	179
Db	631	GAGGGAGCTGTGTGTATCTATGAACCTACCGAGTGGGAACCTTTGGCTTCTTGCCCTA	690
QY	180	AspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMet	199
Db	691	CCAGGAACGAGA---GAAGCCCTGGCAATGTAGTCTGTGGATCAACGGTTCCTGTG	747
QY	200	GlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysValThrIlePhe	219
Db	748	CAATGGGTGCAAGAAATATTTCAGCCTTTGGGGCGCACCCGATGTCACTGCTGTTT	807
QY	220	GlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAspGlyAspAsn	239
Db	808	GGGGAGAGTCCGGTGCAGCCTCCGTGGGCATGCACATACTG-----	849
QY	240	ThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValPro	259
Db	850	TCCCTGCCAGCAGGAGCCTCTTCCACAGGGGTGTCTCCACAGTGGC-----ACACCC	903
QY	260	SerAspPro-----ValAspGlyThrTyrGlyAsnGluIleTyrAspLeuPheVal	276
Db	904	ATGGGCCCTGGGCCACTGTGTAGTGTGTGGAGGCCAGCGCGAGGCCACACTGTGGCC	963
QY	277	SerSerAlaGlyCys-----GlySerAlaSerAspLysLeuAlaCys	290
Db	964	CGCCTTGTGGGTGTCTCCCGAGGTGGCGTGTGGCAATGACACCGAGTGAAGCTGTC	1023
QY	291	LeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeu	310
Db	1024	TTGAGGACAGGCCCGCTCAGGACCTGTGGACACGAGTGGCACCTC-----CTG	1074
QY	311	AlaTyrSerSerLeu---ArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThr	329
Db	1075	CCTCAAGAAAGTATCTCCGATTTCTTCGTGCTGTGTGTGACGCGGACTTCTCTCAGT	1134


```

/codon_start=1
/protein_id="BAC34196.1"
/db_xref="GI:26341068"
/translation="MOTQHTKVYQTHFLMILLCHPFGKSHTEEDFIITKTRVRG
LSPVLGTVTAFLGIPYAPPIGSLRFRKQPLNKPDPIDHNATQVANSYQNILOAF
PGFGSEMNPNLSEGLYLNMTVPKPKNATVMVTIYGGGFGTGTSSLPVYDGK
FLARVERIVSNMVRVGLAFPGNPDAGNMGFLDQOLALQWVORNIARAFCPGN
KSIITFGESAGASVSLHLLCPQSYPLFTFALLESSSNAPAVKHPPEARNRTLTLA
KFTGCKSEMEMIKLRSKDPOEILNERNRPVLPSPSILSINFGPTVDGDFLDMPHIT
LLQGLKVKQAQIILGVNNDKGTAFVYAGFGSKNDLSLITRKEFQGLGNMYPGVSR
LQKAVLFYVVDLWGEQSPVYRDALDVIYDNIICPALEFKKFAELNNAPFFYF
EHRSSKLPPWEMMGHGEIEFVGLGRVNYTRAEBIFSRISMKTKWANFAKYGH
PRTQGNSTWMPVFTSTEQKVLTLNTEKSKYKLRAPQCQFWRLFPKPLEMTCTDID
ETSQEWKAGFHRWSNMNDQNDYTSKESCTAL"

```

ORIGIN

```

Alignment Scores:
Pred. No.: 1.38e-41 Length: 3206
Score: 471.00 Matches: 164
Percent Similarity: 44.57% Conservative: 74
Best Local Similarity: 30.71% Mismatches: 194
Query Match: 16.45% Indels: 102
DB: 11 Gaps: 24

```

US-09-943-857-4 (1-547) x AK050337 (1-3206)

```

Qy 22 GlyAspThrIleThrGlyLeuAsnAlaIlelehenGluAlaPheLeuGlyIleProPhe 41
|||:|||||
Db 441 GGTGGCAGCGTGACT-----GCCCTTCTCGGTATCCCTAT 476

Qy 42 AlaGluProProValGlyAsnLeuArgPheLysAspProValProTyrSerGlySerLeu 61
|||:|||||
Db 477 GCACAACTCTCTCGGTAGCTTAAAGTTCACAAAGCGCAACCTTAAACAAATGGCCT 536

Qy 62 AenGlyGlnLysPheThrSerGlyProSerCysMetGlnGlnAsnProGluGlyThr 81
|||:|||||
Db 537 GACATCCATAATGCCACTCAATATGCAATCTTGT---TATCAGAACATAGCAACAGCC 593

Qy 82 PheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGln 101
|||:|||||
Db 594 TTCCAGAGGCTTCCAGGG-----TCAGAAATGGGAAT 626

Qy 102 AlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgPro---ProGly 120
|||:|||||
Db 627 CCAACACAAACTCGTGAAGACTGCTGTATCTGAATGTTTGGATTCAGTACCGAAG 686

Qy 121 ThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyGlyPheGluIle 140
|||:|||||
Db 687 CCTAAA-----AATGCCACTGTCATGTGTATGATCTATGTGTGGCTTTCAAAC 737

Qy 141 GlySerProThrIlePheProAlaGlnMetValThrLysSerValLeuMetGlyLys 160
|||:|||||
Db 738 GGGACCTCTTCTCTA-----CCTGTTTACGATGGGAAG 770

Qy 161 -----HisIleIleHisValAlaValAsnTyrArgValAlaSerTrp 174
|||:|||||
Db 771 TTTCTAGCTCGTGTGAAGAGTATTGTAGTTTCGATGAATATAGGTAGGTGCTCTA 830

Qy 175 GlyPheLeuAlaGlyAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAsp 194
|||:|||||
Db 831 GGATTCCTAGCTTTTCCGGA---AATCCGATGCTCCAGGAACATGGGTTTGTGTAT 897

Qy 195 GlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSer 214
|||:|||||
Db 888 CACACAGTGGCACTTCAATGGGTCCAAAGAAATATAGCTCTTTTGAGGGAATCTTAA 947

Qy 215 LysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrp 234
|||:|||||
Db 948 AGTATACGATTTTGGAGAAAGTGCAGGGGCAGCTTCACTTACATTTGCTC--- 1004

Qy 235 AsnAspGlyAspAsnThrTyrLysGlyLysPheProLeuPheArgAlaGlyIleMetGlnSer 254
|||:|||||
Db 1005 -----TGCCCCCAAGATTATCTTTGTTTACCAGAGCCATCTTCTGAAGT 1049

```

```

Qy 255 GlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGluIleTyrAspLeu 274
|||:|||||
Db 1050 GGCTCC-----TCTAATGCCCCCTGGCGAGTAAAGCATCCTCAGGAAGCCAGAAACAGA 1103

Qy 275 PheValSerSerAla-----GlyCysGlySerAlaSer-----AspLysLeuAla 289
|||:|||||
Db 1104 ACCTTCACCTTAAATTTACTGGTTGCTCAAGGAAATGAGATGGAGATGATTAAA 1163

Qy 290 CysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPhe 309
|||:|||||
Db 1164 TGCTTCGAATGAAGATCCTCAGGAATCTT-----CGCAATGAAGAGTTCGTCTC 1217

Qy 310 LeuAlaTyrSerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThr 329
|||:|||||
Db 1218 CCCTCTCATTCATCTTATCCATAAATTTGGTCCAAACAGTGGATGGGATTTCTCACC 1277

Qy 330 AspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleIleGly 349
|||:|||||
Db 1278 GATATGCCCCACACACTACTCCAACTAGGAAAGTGAAGAAAGCTCAGATCTTAGTGGGA 1337

Qy 350 AspGlnAsnAspGluGlyThrIlePhe-----GlyLeuSerSerLeu 363
|||:|||||
Db 1338 GTTAAACAAAGATGAAGGACAGCTTCTCTAGTGTACGCTGCTCGGGTTTCAGCAAGAC 1397

Qy 364 Asn-----ValThrThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPheIleHis 381
|||:|||||
Db 1398 AATGATAGCTTATCACAAGGAAGAAATTCAGAAGGTTTAAATATATGATTTCCTCGGA 1457

Qy 382 AlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyrProGlnAspIleThrGlnGly 401
|||:|||||
Db 1458 GTGAGCAGATTGGGCAAGGAGCAGTCTTTTCTACTACGTGAGTGTAGTGTAGTGAGCAG 1517

Qy 402 SerPro-----PheAsp-----ThrGlyValLeuAsnAlaLeuThr 413
|||:|||||
Db 1518 TCACCAAGAGTCTACCGTGACGCTTGGATGATGTTATTGGAGATTACACATCATCTGCG 1577

Qy 414 Pro-----GlnPheLysArgIleSerAlaValLeuGlyAlaPheIleHisAla 431
|||:|||||
Db 1578 CTTGCACTGGAGTTTACCAAGAAATTTGCAGAGCTTGAACAAATGCTTTTTC----- 1631

Qy 432 ArgArgTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSerLysGln 451
|||:|||||
Db 1632 -----TACTTTTCGAACAT-----CGATCTTCCAAA 1658

Qy 452 LeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAspTrp 471
|||:|||||
Db 1659 CTACCTTGGCCGAATGGATGGAGTGCATGCTATGCTATGAAAT-----GAATTT 1709

Qy 472 LeuLeuGly-----SerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAla 488
|||:|||||
Db 1710 GTGTTTGGCTTACCTCTCGGAAAGAGATTAAATTATACGAGAGCTGAGGAATCTTAGT 1769

Qy 489 ThrAspLeu-----AspProAsn-----ThrAla 496
|||:|||||
Db 1770 CGATCCATATGAATACTGGCAAAATTTTGCAAAATATGGACATCCCAATGGACCCAG 1829

Qy 497 GlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGln 510
|||:|||||
Db 1830 GGCAATAGCACAAATGTGGCTCTGTCTTCAACAGTACTGAAACAA 1871

```

RESULT 11

```

AK052200      2026 bp      mRNA      linear      HTC 20-SEP-2003
LOCUS      Mus musculus 13 days embryo heart cDNA, RIKEN full-length enriched
DEFINITION      library, clone:D33006P10 product:butyrylcholinesterase mRNA, full
                insert sequence.
ACCESSION      AK052200
VERSION        AK052200.1 GI:26095058
KEYWORDS      HTC; CAP trapper.
SOURCE        Mus musculus (house mouse)
ORGANISM      Mus musculus
                Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

```


QY	275	pheValserSerAla-----GlyCysGlySerAlaSer-----AspLysLeuAla	289	JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
Db	932	ACCTTGACCTTACGTAAATTTACTGTGCTCAAGGAAATGAGTGGAGTATTA	991	PUBMED	99279253
QY	290	CysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPhe	309	REFERENCE	10349636
Db	992	TGCTCTCGAAGTAAAGATCTCTAGGAATTCCT-----CGCATGAAAGTTCGTCTC	1045	AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
QY	310	LeuAlaTyrsSerSerLeuArgLeuSerTyrsLeuProArgProAspGlyLysAsnIleThr	329	TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Db	1046	CCTCTGATTCATCTATCCATAAATTTGGTCCACAGTGGATGCGGATTTCTCACC	1105	JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
QY	330	AspAspMetTyrsLysLeuAlaArgAspGlyLysTyrsAlaSerValProValIleIleGly	349	PUBMED	20499374
Db	1106	GATATGCCCCACACACTACTCTCACTAGGAAAGTGAAGCTCAGATCTTAGTGGGA	1165	REFERENCE	11042159
QY	350	AspGlnAsnAspGluGlyThrIlePhe-----GlyLeuSerSerLeu	363	AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kiteunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Watanabe, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, K., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
Db	1166	GTTAACAAAGATGAAGGACAGCTTCTCTAGTGACGTCTCGGTTTCAGCAAGAC	1225	TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
QY	364	Asn-----ValThrThrAsnAlaGlnAlaArgAlaTyrsPheLysGlnSerPheIleHis	381	JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
Db	1226	AATGATAGCTTATCACACAGGAGGAATTTCAAGAGGTTTAAATATGTATTTCCCTGGA	1285	PUBMED	20530913
QY	382	AlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyrsProGlnAspIleThrGlnGly	401	REFERENCE	11076861
Db	1286	GTGAGCAGATTGGGCAAGAGCAGTCTCTTTCTACTAGCTGGAGTGGTTAGTGAGCAG	1345	AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
QY	402	SerPro-----PheAsp-----ThrGlyValLeuAsnAlaLeuThr	413	TITLE	Functional annotation of a full-length mouse cDNA collection
Db	1346	TCACAGAAGCTACCGTGACGCTTTGGATGATGTATTGGAGATTACACATCATCTGC	1405	JOURNAL	Nature 409, 685-690 (2001)
QY	414	Pro-----GlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHisAla	431	REFERENCE	5
Db	1406	CTGCACCTGGAGTTTACCAGAGAAATTTGCAGAGCTTGAACAACAATGTTTTTC	1459	AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
QY	432	ArgArgTyrsPheLeuAsnHisPheGlnGlyGlyThrLysTyrsSerPheLeuSerLysGln	451	TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Db	1460	-----TACTTTTTCGAACAT-----CGATCTTCCAAA	1486	JOURNAL	Nature 420, 563-573 (2002)
QY	452	LeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAspTyrs	471	REFERENCE	6
Db	1487	CTACCTTGCCCGAATGGATGGAGTGATGCTGCTATGAATT-----GAATTT	1537	AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanegaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanakawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
QY	472	LeuLeuGly-----SerGlySerValIleTyrsAsnAsnAlaPheIleAlaPheAla	488	TITLE	Direct Submission
Db	1538	GTGTTGGCTTACCTCTGGAGAGAGAGTAAATATATACGAGAGCTGAGGAATCTTAGT	1597	JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
QY	489	ThrAspLeu-----AspProAsn-----ThrAla	496	COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken
Db	1598	CGATCCATAATGAAAACTTGGCAAAATTTGGAAAAATATGGACATCCCATGGACCCAG	1657	REFERENCE	Encyclopedia Project of Genome Exploration Research Group in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues
QY	497	GlyLeuLeuValAsnTrpProLysTyrsThrSerSerSerGln	510	FEATURES	Please visit our web site for further details.
Db	1658	GGCAATAGCAATATGGCTGCTCTTCACAGTACTGACAA	1699	source	URL:http://genome.gsc.riken.go.jp/ URL:http://fantom.gsc.riken.go.jp/ Location/Qualifiers 1. .2177 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM DB:9630050A01" /db_xref="MGI:2400667" /db_xref="taxon:10090" /clone="9630050A01"
RESULT 12	AK036259	2177 bp	mRNA	linear	HTC 19-SEP-2003
LOCUS	AK036259	Mus musculus 16 days neonate cerebellum cDNA, RIKEN full-length			
DEFINITION		enriched library, clone:9630050A01 product:acetylcholinesterase, full insert sequence.			
ACCESSION	AK036259				
VERSION	AK036259.1	GI:26085119			
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	1				
AUTHORS	Carninci, P. and Hayashizaki, Y.				
TITLE	High-efficiency full-length cDNA cloning				

/tissue_type="cerebellum"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="16 days neonate"
 111..1955
 /note="acetylcholinesterase (MGI:87876, GB|NM_009599,
 evidence: BLASTN, 99%, match=1845)
 putative."
 2161..2166
 /note="putative"
 2177
 /note="putative"

misc_feature

polyA_signal

polyA_site

ORIGIN

Alignment Scores:

Pred. No.: 8,53e-41 Length: 2177
 Score: 461.50 Matches: 162
 Percent Similarity: 44.49% Conservative: 80
 Best Local Similarity: 29.78% Mismatches: 169
 Query Match: 16.11% Indels: 134
 DB: 11 Gaps: 26

US-09-943-857-4 (1-547) x AK036259 (1-2177)

QY 35 AlaPheLeuGlyLeuPheAlaGluProValGlyAsnLeuArgPheLysaspPro 54
 DB 294 GCTTTTCTGGGATCCCTTTCAGAGCCACCTGTGGCTCAGTAGATTTATGCCACCA 353
 QY 55 Val-----ProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyPro 71
 DB 354 GAGCCCAAGCGCCCTGGTCAGAGGTGTGGATGCT-----ACCACCTCCAAAT 404
 QY 72 SerCysMetGln-----GlnAsnProGluGlyThrPheGluGluAsn 35
 DB 405 GTCTGTACCACTAGCTGGACACCCCTGTACCTGGGTTTGGGTACTCAGATGTGGAC 464
 QY 86 LeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuPro 105
 DB 465 CCCAACCGAGAGTTG-----GATGTGTATGACGGCCGTTTCTGTGCCACAGGTT 526
 QY 106 GlnSerGluAspCysLeuThrIleAsnValValArgProGlyThrLysAlaGlyAla 125
 DB 480 ---AGTGAAGACTGCTGTATCTTATGTGTGGACACCATACCCAGACCTGCTTCTCC 536
 QY 126 AsnLeuProValMetLeuThrIlePheGlyGlyPheGluIleGlySerProThrIle 145
 DB 537 ACA---CCTGTCTCATCTGGATCTATGGGTGGTGTTCACAGCGAGCGGCTCCCTG 593
 QY 146 PheProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIle----- 162
 DB 594 -----GATGTGTATGACGGCCGTTTCTGTGCCACAGGTT 526
 QY 163 -----IleHisValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGly 179
 DB 627 GAGGGAGTGTGTGTATCTATGACTACCGAGTGGGAACCTTTGGCTTCTTGCCCTTA 686
 QY 180 AspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMet 199
 DB 687 CCAGGAAGCAGA---GAAGCCCTGGCAATGAGTCTGTGATCAACGGCTTGCCTTG 743
 QY 200 GlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysValThrIlePhe 219
 DB 744 CAATGGGTGCAGAAAAATATTGCACGCTTTGGGGCGGACCCGATGTCAGTCACTCTGTT 803
 QY 220 GlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAspGlyAspAsn 239
 DB 804 GGGGAGATGGGGTGGAGCTCCGTGGCATGCACATACCTG----- 345
 QY 240 ThrTrpLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValPro 259
 DB 846 TCCCTGCCAGCAGGAGGCTCTTCCACAGGGGTGCTCCAGAGTGGC-----ACACCC 899
 QY 260 SerAspPro-----ValAspGlyThrTyrGlyAsnGluIleTyrAspLeuPheVal 276
 DB 940 AATGGCCCTGGGCCACTGTGAGTGTCTGAGAGGCCAGCGCAGGCGCACACTGTCTGGCC 959
 QY 277 SerSerAlaGlyCys-----GlySerAlaSerAspLysLeuAlaCys 290
 DB 960 CGCTTGTGGCTGTCCCCAGGTGGCGTGTGGCAATGACACCGAGCTGATAGCTGC 1019
 QY 291 LeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeu 310
 DB 1020 TTGAGGACAAGGCCCTCAGGACCTGGTGGACCCAGAGTGGCAGCTC-----CTG 1070
 QY 311 AlaTyrSerSerLeu---ArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThr 329
 DB 1071 CCTCAAGAAAGTATCTTCGATTTTCTTCGTGCTGTGTGACACGGGACTTCTCTCAGT 1130
 QY 330 AspAspMetTyrLysLeuValArgaspGlyLysTyrAla-SerValProValIleIleG 349
 DB 1131 GACACACCGAGGCTCTCATCAATACTGGAGATTTTCAAGACCTGCAGCTGTGTGGG 1190
 QY 349 yAspGlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnAl 369
 DB 1191 TGTGTGAAGGACGAGGCTCTACTTT-----CTGTTACGGGGTCT 1232
 QY 369 aGlnAlaArgAlaTyrPheLysGlnSerPheIle----- 380
 DB 1233 CCAGGCTTCAGCAAGAC-AATCAATCTCTCATCAGCGCGGCGCAGTTCTCTGGTGGGT 1291
 QY 381 -----HisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyr--- 394
 DB 1292 GCGGATCGGTATCCCCAAGCAAGTGACCTGGCAGCGGAGGCTGTGCTCTCATTTACAC 1351
 QY 395 -----ProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeu--- 409
 DB 1352 AGACTGGCTGACCTGAGACCTTACTCACCTGAGAGATGCCATGAGTGCAGTGGTAGG 1411
 QY 410 -----AsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAl 427
 DB 1412 CGACCACAACGTTGTGTGCCT-----GTGGCCAGCTGGCTGGCGACTGGC 1459
 QY 427 aPheIleHisAlaArgTyrPheLeuAsnHisPheGlnGlyGlyThrLysTyrSerPh 447
 DB 1460 T-----GCCAAGGGCGCCGGTCTATGCTTA 1486
 QY 447 eLeuSerLysGlnLeuSerGlyLeu-----ProIle---MetGlyThrPheHisAl 463
 DB 1487 CATCTTTGAACACCGTGCCTCCACACTGACTGGCCCTCTCGATGGGGTGCCTCATGG 1546
 QY 463 aAsnAspIleValTrpGlnAspTyrLeuLeuGly-----SerGlySerValIleTyr 480
 DB 1547 CTATGAATC-----GAGTTTCATCTTTGGGTCTCCCTGGATCCCTCGTGAACCTA 1597
 QY 480 rAsnAsnAlaPheIleAlaPheAlaThrAspLeu----- 491
 DB 1598 CACACGAGGAGAGGATCTTTGCTCAGCGACTTATGAATACTGACCAATTTTGGCCG 1657
 QY 492 -----AspProAsnThrAla-----GlyLeuLeuValAsnTrpProLysTyrThrSe 507
 DB 1658 CACAGGGAGCCCAATGACCCCTCGAGACTCCAAATCTCCACAGTGGCCACCGTACACCC 1717
 QY 507 rSerSerGln 510
 DB 1718 TGCCGCGCAG 1727
 RESULT 13
 AW790714
 LOCUS
 DEFINITION D00142-R Lambda Zap, Striatogene Blumeria graminis f. sp. hordei
 cDNA clone D00142 similar to lipase 2 precursor, mRNA sequence.
 ACCESSION AW790714
 VERSION AW790714.1 GI:13902311
 KEYWORDS EST.
 SOURCE Blumeria graminis f. sp. hordei
 ORGANISM Blumeria graminis f. sp. hordei
 Eukaryota; Fungi; Ascomycota; Perizomycotina; Leotiomycetes;


```

Db      6  TGGAAACCTATTGACTGCTAAGGATGCTAACCTACGATACACGACAGCCCTCTGTCCGA 55
Qy      249 AlaGlyIleMetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTyGly 268
Db      66  GCTGCATATAGGACGAGGAGTGTCTATCCACAGACCCAGCTGCTGACCTAAAGCC 125
Qy      269 AsnGluIleTyAspLeuPheValSerSerAlaGlyCysGlySerAlaSerAspLysLeu 288
Db      126 CAGAACATTTTAACTGCTGCTGCTGCTGAGGCTGTGAAGGCGCTGCTGACAAAGATC 185
Qy      289 AlaCysLeuArgSerAlaSerAspThrLeuLeuAspAlaThrAsnAsnThrProGly 308
Db      186 GACTGCCTCCAGCTGCTGCTTACGATCAGTTCCTCCAGGCCAGTACATCCGTACCTGGT 245
Qy      309 PheLeuAlaTySerSerLeuArgLeuSerTyLeuProArgProAspGlyLysAsnIle 328
Db      246 CTATTCGATTACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 305
Qy      329 ThrAspMetTyLeuValArgAspGlyLysTyAlaSerValProValIleIle 348
Db      306 ACCGATTCCGAGATACCTCTCGAGGAGGAGGATGCAACGTTCCCTACATCGTC 365
Qy      349 GlyAspGlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrAsn 368
Db      366 GGTGATCAAGAAGACGAGGAGTACCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 425
Qy      369 AlaGlnAlaArgAlaTyPheLysGln---SerPheIleHisAlaSerAspAlaGluIle 387
Db      426 CAGGACATCATCGACTACTTCAAGCTTTTACTCTCTCACGCTAGTATCAATTACTA 485
Qy      388 AspThrLeuMetAlaAlaTyProGlnAspIleThrGlnGlySerProPheAspThrGly 407
Db      486 TCCGACCTTGTAACTTGTACCGAGATGACCCCGCGCTGGATCGGCATACACACTGGA 545
Qy      408 ValLeuAlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeu 426
Db      546 ACTCGTAACAACATATACCTCAATACAGAGACTTGGCTATTATCGTGACATG 602

RESULT 15
LOCUS   AY407736
DEFINITION Homo sapiens Ache gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.
ACCESSION AY407736
VERSION   1
KEYWORDS GSS.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferrera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Gargill,M.
INFERRED nonneutral evolution from human-chimp-mouse orthologous gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferrera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Gargill,M.
TITLE Direct Substitution
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering them based on alignment.
FEATURES
source location/Qualifiers
1..1723
/organism="Homo sapiens"
/mol_type="genomic DNA"

```

```

/db_xref="taxon:9606"
<1..>1723
/feature="ACHE"
/locus_tag="HCM2990"

```

ORIGIN

```

Alignment Scores:
Pred. No.: 6.89e-40 Length: 1723
Score: 452.00 Matches: 156
Percent Similarity: 44.59% Conservative: 83
Best Local Similarity: 29.10% Mismatches: 179
Query Match: 15.78% Indels: 118
DB: 29 Gaps: 26

US-09-943-857-4 (1-547) x AY407736 (1-1723)

Qy      35  AlapheLeuGlyIleProPheAlaGluProValGlyAsnLeuArgPheLysAspPro 54
Db      184 GCITTCCTGGGCATCCCTTTGGGAGCCACCCATGGGACCCCTGCTCTTCTGCCACCG 243
Qy      55  -----ValProTySerGlySerLeuAsnGlyGlnLysPheThrSerTyGlyPro 71
Db      244 GAGCCCAAGCAGCGCTTGGTCAGGGGTGTAGACGCTACAACTTCCAGAGTGC----- 297
Qy      72  SerCysMetGln-----GlnAsnProGluGlyThrPheGluGluAsn 85
Db      298 ---TGCTACCAATATGTGGACACCTATACCCAGGTTTTCAGGGCCACCGAGATGTGGAAC 354
Qy      86  LeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuPro 105
Db      355 CCCAACCGTGAGCTG-----GAGGTGTAGATGGCCGCTTGTGTACAGGCC 516
Qy      106 GlnSerGluAspCysLeuThrIleAsnValValArgProGlyThrLysAlaGlyAla 125
Db      370 ---AGCGAGGACTGCTGTCCATGAATACCGGGTGGGAGCCTTTGGCTTCTCGGCCCTG 423
Qy      126 AsnLeuProValMetLeuTrpIlePheGlyGlyGlyPheGluIleGlySerProThrIle 145
Db      424 CCCACCCCTGCTGCTGTGGATCTATGGGGTGGCTTCTACAGTGGGGCTCTCTCTCTG 483
Qy      146 PheProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIleIleHis--- 164
Db      484 -----GAGGTGTAGATGGCCGCTTGTGTACAGGCC 516
Qy      165 -----ValAlaValAsnTyArgValAlaSerTrpGlyPheLeuAlaGly 179
Db      517 GAGAGGACTGTGCTGTCCATGAATACCGGGTGGGAGCCTTTGGCTTCTCGGCCCTG 576
Qy      180 AspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMet 199
Db      577 CCGGGGAGCCGA---GAGGCCCGGGCAATGTGGGTCTCTGGATCAGAGGCTGGCCCTG 633
Qy      200 GlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysValThrIlePhe 219
Db      634 CAGTGGGTGCAGGAGAGCTGCGAGCTTCGGGGTGTACCCAGCATCATGACGCTGTT 693
Qy      220 GlyLysAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAspGlyAspAsn 239
Db      694 GGGGAGAGCGCGGAGCGCCCTCGTGGGCATGCACCTGCTG-----AsnGluIle 735
Qy      240 ThrTyLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValPro 259
Db      736 TCCCGCCAGCCGGGCGCTTTCACAGGGCCGTGCTGCAGAGCGCGTGC-----CCC 789
Qy      260 SerAspProValAspGlyThrTyGly-----AsnGluIle 271
Db      790 AATGGACCTCTG---GCCAGGTGGGATCGGAGAGCGCCGCTGCAGAGGCCACGACGT 846
Qy      272 TyrAspLeuPheValSerSerAlaGlyCys-----GlySerAlaSer 285
Db      847 GCCACGCTT-----GTGGGCTGCTCCAGGGCGGCACTGGTGGGAATGACACA 894
Qy      286 AspLysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsn 305

```

Fri Aug 6 10:53:13 2004

```

Db      895  GAGCTGGTAGCTGCTTGGACACGACCGAGGTCCTGGTG-----AACCAC 945
Qy      306  ThrProGlyPheLeuAlaTyrSerSerLeu---ArgLeuSerTyrLeuProArgProAsp 324
Db      946  GAATGGCAGCTGCTGCTCAAGAAAGCGTCTTCGGTTCTCTCGTGGCTGTGGTAGAT 1005
Qy      325  GlyLysAsnIleThrAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerVal 344
Db      1006  GGAGACTTCTCAGTGACACCCAGAGGCCCTCATCAGCGGAGACTTCACGGCCTG 1065
Qy      345  ProValIleIleGlyAspGlnAsnAspGluGlyThrIlePhe----- 358
Db      1066  CAGGTGCTGGTGGTGTGGTGAAGATGAGGGCTCGTATTTCTGGTTTACGGGGCCCCA 1125
Qy      359  GlyLeuSerSerLeuAsn-----ValThrAsnAlaGlnAlaAlaGlyTyrPheLys 376
Db      1126  GGCTTCAGCAAGACACAGAGTCTCTCATCAGCGGGCCGAGTTCTTGGCCGGGTGCGG 1185
Qy      377  GlnSerPheIleHisAlaSerAlaGluIleAspThrLeuMetAlaAlaTyr----- 394
Db      1186  GTCGGGTTCCCGCAGGTAAGTACCTGGCAGCCGAGGCTGTGGTCTCTGTCATTACAGAC 1245
Qy      395  -----ProGlnAsp-----IleThrGlnGlySerProPheAspThrGlyVal 408
Db      1246  TGGCTGCATCCCGAGGACCCGCGCAGCTGAGGGAGCCCTGAGCGATGTGGGGCGAC 1305
Qy      409  LeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPhe 428
Db      1306  CACAAATGCTGTGTCCTCC-----GTGGCCAGCTGGCTGGGGCGACTGGCTGCC 1353
Qy      429  IleHisAlaArg-----ArgTyrPheLeuAsnHisPheGlnGly----- 441
Db      1354  CAGGGTGCCTGGGTCTAGCCCTACGCTCTTGAACACCGGTCTTCCAGCTCTCTCTGGCCC 1413
Qy      442  -----GlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeuProIle--- 457
Db      1414  CTGTGGATGGGGGTGCCCCCAGCGCTACGAGATCGAGTTCATCTTTGGGATCCCCCTGGAC 1473
Qy      458  ---MetGlyThrPheHisAlaAsnAspIleValTrpGlnAspTyrTrpLeuGlySerGly 476
Db      1474  CCTCTCGAAACTACACGGCAGAGGAGAAATCTTCGCCCGCAGCGACTGATG----- 1524
Qy      477  SerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAla 496
Db      1525  -----CGATACTGGGCCCACTTTGCCCGCACAGGGGATCCCAATGAGCCC 1569
Qy      497  -----GlyLeuValAsnTrpProLysTyrThrSerSerSerGln 510
Db      1570  CGAGACCCCAAGGCCCCACAAATGGCCCCCGTACACGGCGGGGCTCAG 1617

```

Search completed: August 5, 2004, 21:18:56
Job time : 2731 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 10:43:29 ; Search time 13 Seconds
(without alignments)
2190.952 Million cell updates/sec

Title: US-09-943-857-4
Perfect score: 2864
Sequence: 1 SMNSRGAGRLGSVPTAKLA.....DNFTAGYDALMTNPSSFFV 547

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2779	97.0	549	1 LIP3 CANRU	P32947 candida rug
2	2513	87.7	549	1 LIP1 CANRU	P20261 candida rug
3	2431	84.9	549	1 LIP2 CANRU	P32949 candida rug
4	2366	82.6	549	1 LIP4 CANRU	P32948 candida rug
5	2314	80.8	548	1 LIP2 CANRU	P32946 candida rug
6	1092	38.1	563	1 LIP1 GEOCN	P17573 geotrichum
7	1078	37.6	563	1 LIP2 GEOCN	P22394 geotrichum
8	491.5	17.2	586	1 ACES_TORCA	P04058 torpedo cal
9	491	17.1	633	1 ACES_ELEEL	O42275 electrophor
10	489	17.1	590	1 ACES_TORMA	P07692 torpedo mar
11	482	16.8	574	1 CHLE_HORSE	P81908 equus cabal
12	481	16.8	614	1 ACES_MOUSE	P21836 mus musculu
13	476	16.6	603	1 CHLE_MOUSE	O03311 mus musculu
14	472	16.5	629	1 ACES_LEPDE	O27577 leptinotars
15	471	16.4	613	1 ACES_BOVIN	P23795 bos taurus
16	470	16.4	611	1 ACES_FELCA	O62763 felis silve
17	469	16.4	614	1 ACES_RAT	P37136 rattus norv
18	462	16.1	606	1 ACES_BUNFA	Q92035 bungarus fa
19	459	16.0	581	1 CHLE_RABIT	P21927 oryctolagus
20	459	16.0	602	1 CHLE_FELCA	O62760 felis silve
21	458.5	16.0	634	1 ACES_BRARE	Q98de3 brachydanio
22	458	16.0	584	1 ACES_RABIT	Q29499 oryctolagus
23	455	15.9	602	1 CHLE_PANTT	O62761 panthera t
24	453.5	15.8	737	1 ACES1 ANOGA	Q869c3 anopheles g
25	452	15.8	614	1 ACES_HUMAN	P22303 homo sapien
26	451	15.7	602	1 CHLE_HUMAN	P06276 homo sapien
27	445.5	15.6	702	1 ACES_CULPI	Q869c8 cullex pipie
28	437	15.3	620	1 ACES1_CABR	Q27459 caenorhabdi
29	431.5	15.1	489	1 PNBA_BACSD	P37967 bacillus su
30	427.5	14.9	620	1 ACES1 CAEEL	P38433 caenorhabdi
31	423	14.8	532	1 EST2_RABIT	P14943 oryctolagus
32	413	14.4	549	1 EST1_RAT	P10959 rattus norv
33	411.5	14.4	597	1 CEL_BOVIN	P30122 bos taurus

34	411	14.4	664	1 ACES_ANOST	P56161 anopheles s
35	408.5	14.3	561	1 EST1_MESAU	Q64419 mesocricetu
36	408	14.2	565	1 EST10_RAT	P16303 rattus norv
37	402	14.0	561	1 EST4_RAT	Q64573 rattus norv
38	401	14.0	554	1 ESTN_MOUSE	P23953 mus musculu
39	401	14.0	649	1 ACES_DROME	P07140 drosophila
40	396	13.8	557	1 SASB_ANAPL	Q04791 anas platyr
41	396	13.8	767	1 ACES_CHICK	P36196 gallus gall
42	394	13.8	742	1 CEL_HUMAN	P19835 homo sapien
43	393.5	13.7	561	1 EST5_RAT	Q63010 rattus norv
44	392	13.7	561	1 EST3_RAT	Q63108 rattus norv
45	390	13.6	599	1 CEL_MOUSE	Q64285 mus musculu

ALIGNMENTS

RESULT 1					
LIP3 CANRU	ID	LIP3 CANRU	STANDARD;	PRT;	549 AA.
AC	P32947;				
DT	01-OCT-1993	(Rel. 27, Created)			
DT	01-OCT-1993	(Rel. 27, Last sequence update)			
DT	15-MAR-2004	(Rel. 43, Last annotation update)			
DE	Lipase 3 precursor (EC 3.1.1.3) (Cholesterol esterase).				
GN	Lip3				
OS	Candida rugosa (Yeast) (Candida cylindracea).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OC	Saccharomycetales; mitosporic Saccharomycetales; Candida.				
OX	NCBI_TaxID=5481;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=ATCC 14830;				
RX	MEDLINE=93178975; PubMed=8440480;				
RA	Lotti M., Grandori R., Fusetti F., Longhi S., Brocca S.,				
RA	Tramontano A., Alberghina L.;				
RT	"Cloning and analysis of Candida cylindracea lipase sequences.";				
RL	Gene 124:45-55(1993).				
RN	[2]				
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).				
RX	MEDLINE=95308044; PubMed=7788294;				
RA	Ghosh D., Wawrzak Z., Pletnev V.Z., Li N., Kaiser R., Pangborn W.,				
RA	Joernvall H., Erman M., Duax W.L.;				
RT	"Structure of uncomplexed and linoleate-bound Candida cylindracea				
RT	cholesterol esterase.";				
RL	Structure 3:279-288(1995).				
RN	[3]				
RP	REVIEW.				
RX	MEDLINE=98451816; PubMed=9778794;				
RA	Benjamin S., Pandey A.;				
RT	"Candida rugosa lipases: molecular biology and versatility in				
RT	biotechnology.";				
RL	Yeast 14:1069-1087(1998).				
CC	-!- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a				
CC	fatty acid anion.				
CC	-!- SIMILARITY: Belongs to the type-B carboxylesterase/lipase family.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
CC	EMBL; X66006; CAA46805.1; -				
DR	PIR; JN0551; JN0551.				
DR	PDB; ICLE; 08-MAR-96.				
DR	PDB; LILF; 07-JAN-03.				
DR	InterPro; IPR002018; CarbesteraseB.				
DR	InterPro; IPR000379; Ser_estr.				
DR	Pfam; PF00135; Coesterase_1.				
DR	PROSITE; PS00122; CARBOXYLESTERASE_B_1; 1.				

DR PROSITE; PS00941; CARBOXYL ESTERASE_B_2; 1.
KW Cholesterol metabolism; Lipid degradation; Hydrolase; Signal;
KW Glycoprotein; Multigene family; 3D-structure.
FT SIGNAL 1 15
FT CHAIN 16 549 LIPASE 3.
FT ACT SITE 224 224 BY SIMILARITY.
FT ACT SITE 356 356 BY SIMILARITY.
FT ACT SITE 464 464 BY SIMILARITY.
FT DISULFID 75 112 BY SIMILARITY.
FT DISULFID 283 292 BY SIMILARITY.
FT CARBOHYD 329 329 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 18 20
FT TURN 22 23
FT STRAND 26 28
FT STRAND 30 31
FT STRAND 36 42
FT STRAND 44 44
FT HELIX 49 51
FT TURN 52 53
FT STRAND 58 58
FT TURN 64 65
FT TURN 67 67
FT STRAND 69 69
FT STRAND 74 74
FT TURN 80 81
FT HELIX 88 98
FT HELIX 100 105
FT STRAND 108 108
FT STRAND 114 119
FT TURN 121 122
FT TURN 125 126
FT STRAND 129 135
FT TURN 139 141
FT HELIX 145 147
FT HELIX 151 159
FT TURN 160 161
FT STRAND 165 169
FT STRAND 174 178
FT HELIX 182 187
FT TURN 188 188
FT HELIX 192 207
FT TURN 208 211
FT TURN 212 212
FT STRAND 213 223
FT TURN 224 224
FT HELIX 225 235
FT STRAND 236 239
FT TURN 242 243
FT STRAND 244 245
FT STRAND 246 247
FT STRAND 251 255
FT TURN 265 266
FT HELIX 268 281
FT TURN 282 282
FT TURN 284 284
FT HELIX 289 294
FT TURN 295 295
FT HELIX 298 305
FT TURN 306 307
FT TURN 311 312
FT TURN 314 317
FT HELIX 334 339
FT TURN 340 341
FT STRAND 348 353
FT TURN 354 354
FT STRAND 355 355
FT HELIX 356 358
FT TURN 359 359
FT HELIX 360 364
FT TURN 365 366
FT HELIX 370 380
FT TURN 382 383

FT	HELIX	386	395
FT	TURN	400	402
FT	TURN	406	407
FT	TURN	409	412
FT	STRAND	413	413
FT	HELIX	418	429
FT	TURN	430	430
FT	HELIX	431	440
FT	STRAND	446	451
FT	TURN	453	456
FT	TURN	458	460
FT	STRAND	463	463
FT	TURN	466	472
FT	HELIX	473	473
FT	TURN	477	478
FT	HELIX	479	482
FT	TURN	483	483
FT	HELIX	484	492
FT	STRAND	495	498
FT	TURN	507	507
FT	TURN	510	511
FT	STRAND	517	520
FT	STRAND	525	528
FT	HELIX	534	541
FT	HELIX	544	547
SQ	SEQUENCE	549 AA;	58754 MW; C5F95C31422975DA CRC64;

Query Match 97.0%; Score 2779; DB 1; Length 549;
Best Local Similarity 98.5%; Pred. No. 1.6e-191;
Matches 529; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY	11	LGSVPTAKLANGDITITGLNAIINEAFIPIPAEPPVGNLRKDPVPYSGSLNGOKFTSYG	70
DB	13	VAAAPTAKLANGDITITGLNAIINEAFIPIPAEPPVGNLRKDPVPYSGSLNGOKFTSYG	72
QY	71	PSCMQQNPETFEENLGKTLALVMQSKVFOAVLPQSEDCITINVRPPGKAGANLPM	130
DB	73	PSCMQQNPETFEENLGKTLALVMQSKVFOAVLPQSEDCITINVRPPGKAGANLPM	132
QY	131	LWIFGGFEIGSPITFFPAQWTVKSLMKGKIIHVAVNYRVASWGFAGDDIKAECSGNA	190
DB	133	LWIFGGFEIGSPITFFPAQWTVKSLMKGKIIHVAVNYRVASWGFAGDDIKAECSGNA	192
QY	191	GLKQRLGQWVADNIAGFGDPPSKVTIFGESAGSMVLCHLIWNGDNTYKGPLFRAG	250
DB	193	GLKQRLGQWVADNIAGFGDPPSKVTIFGESAGSMVLCHLIWNGDNTYKGPLFRAG	252
QY	251	IMQSGAMVPSDPVDGTGNEIYDLFVSSAGCGSSADKLCLRSASDITLLDATNTPGFL	310
DB	253	IMQSGAMVPSDPVDGTGNEIYDLFVSSAGCGSSADKLCLRSASDITLLDATNTPGFL	312
QY	311	AYSSLRLSYLPRPDGKNIITDDMYKLVRDGKYASVPVIGDQNDGTIFGLSSLNVTNAQ	370
DB	313	AYSSLRLSYLPRPDGKNIITDDMYKLVRDGKYASVPVIGDQNDGTIFGLSSLNVTNAQ	372
QY	371	ARAYFKQSFTHASDAEIDTLMAAYPODITQGSPPFTGVNALTPOQFKRISAVIGDLAFIH	430
DB	373	ARAYFKQSFTHASDAEIDTLMAAYPODITQGSPPFTGVNALTPOQFKRISAVIGDLAFIH	432
QY	431	ARRYFLNHFQGGTKYSFLSKQSLGFLPMGTFFHANDIIVWQDYLIGSGSVIYNNAFIATD	490
DB	433	ARRYFLNHFQGGTKYSFLSKQSLGFLPMGTFFHANDIIVWQDYLIGSGSVIYNNAFIATD	492
QY	491	LDENTAGLLVNWPKYTSSSQSGNNLMWINALGLYTGKDNFRTAGYDALTNPSSFFV	547
DB	493	LDENTAGLLVNWPKYTSSSQSGNNLMWINALGLYTGKDNFRTAGYDALTNPSSFFV	549

RESULT 2
LIP1 CANRU
ID LIP1 CANRU
AC P20261;

STANDARD; PRT; 549 AA.

01-FEB-1991 (Rel. 17, Created)
 01-OCT-1993 (Rel. 27, Last sequence update)
 28-FEB-2003 (Rel. 41, Last annotation update)
 Lipase 1 precursor (BC 3.1.1.3).
 LIP1.
 GN
 OS Candida rugosa (Yeast) (Candida cylindracea).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 OC NCBI_TaxID=5481;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 14830;
 RA Longhi S., Fusetti F., Grandori R., Lotti M., Vanoni M.,
 RA Alberghina L.;
 RT "Cloning and nucleotide sequences of two lipase genes from Candida
 RT cylindracea.";
 RL Biochim. Biophys. Acta 1131:227-232(1992).
 RN [2]
 RP SEQUENCE OF 12-549 FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=ATCC 14830 / MS-5;
 RX MEDLINE=89384874; PubMed=2506450;
 RA Kawaguchi Y., Honda H., Taniuchi-Morimura J., Iwasaki S.;
 RT "The codon CUG is read as serine in an asporogenic yeast Candida
 RT cylindracea.";
 RL Nature 341:164-166(1989).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.06 ANGSTROMS).
 RX MEDLINE=93286131; PubMed=8509417;
 RA Grochulski P., Li Y., Schrag J.D., Bouthillier F., Smith P.,
 RA Harrison D., Rubin B., Cygler M.;
 RT "Insights into interfacial activation from an open structure of
 RT Candida rugosa lipase.";
 RL J. Biol. Chem. 268:12843-12847(1993).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.05 ANGSTROMS).
 RX MEDLINE=94190867; PubMed=8142345;
 RA Grochulski P., Bouthillier F., Kazlauskas R.J., Serregei A.N.,
 RA Schrag J.D., Ziomek E., Cygler M.;
 RT "Analogues of reaction intermediates identify a unique substrate
 RT binding site in Candida rugosa lipase.";
 RL Biochemistry 33:3494-3500(1994).
 RN [5]
 RP REVIEW.
 RX MEDLINE=98451816; PubMed=9778794;
 RA Benjamin S., Pandey A.;
 RT "Candida rugosa lipases: molecular biology and versatility in
 RT biotechnology.";
 RL Yeast 14:1069-1087(1998).
 CC -!- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a
 CC fatty acid anion.
 CC -!- SIMILARITY: Belongs to the type-B carboxylesterase/lipase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X64703; CAA45957.1; -;
 DR EMBL: X16712; CAA34684.1; -;
 DR PIR: S05684; S05684.
 DR PIR: S23448; S23448.
 DR PDB: 1LCP; 31-JAN-94.
 DR PDB: 1LPM; 20-APR-95.
 DR PDB: 1LPN; 20-APR-95.
 DR PDB: 1LPO; 20-APR-95.
 DR PDB: 1LPP; 20-APR-95.
 DR PDB: 1LPS; 08-MAR-95.
 DR PDB: 1TRH; 31-JAN-94.
 DR InterPro: IPR002018; CarbesteraseB.

DR InterPro: IPR000379; Ser estrs.
 DR Pfam: PF00135; Coesterase; 1
 DR PROSITE: PS00122; CARBOXYLESTERASE_B_1; 1.
 DR PROSITE: PS00941; CARBOXYLESTERASE_B_2; 1.
 KW Hydrolase; Lipid degradation; Signal; Glycoprotein; Multigene family;
 KW 3D-structure.
 FT SIGNAL 1 15
 FT CHAIN 16 549
 FT ACT_SITE 224 224
 FT ACT_SITE 356 356
 FT ACT_SITE 464 464
 FT DISULFID 75 112
 FT DISULFID 283 292
 FT CARBOHYD 329 329
 FT CARBOHYD 366 366
 FT VARIANT 398 398
 FT STRAND 18 20
 FT TURN 22 23
 FT STRAND 26 28
 FT STRAND 30 31
 FT STRAND 36 42
 FT STRAND 44 44
 FT HELIX 49 51
 FT TURN 52 53
 FT STRAND 58 58
 FT TURN 64 65
 FT STRAND 67 67
 FT STRAND 69 69
 FT STRAND 74 74
 FT TURN 80 81
 FT HELIX 88 97
 FT TURN 98 98
 FT HELIX 100 105
 FT STRAND 108 108
 FT TURN 121 122
 FT TURN 125 126
 FT STRAND 129 135
 FT TURN 139 141
 FT TURN 146 147
 FT HELIX 151 159
 FT TURN 160 161
 FT STRAND 165 169
 FT HELIX 174 178
 FT HELIX 182 187
 FT TURN 188 188
 FT TURN 190 191
 FT HELIX 192 207
 FT HELIX 208 211
 FT TURN 212 212
 FT STRAND 213 223
 FT TURN 224 224
 FT HELIX 225 235
 FT STRAND 236 239
 FT STRAND 242 243
 FT TURN 244 245
 FT STRAND 246 247
 FT STRAND 251 255
 FT TURN 265 266
 FT HELIX 268 281
 FT TURN 282 282
 FT TURN 284 285
 FT HELIX 289 295
 FT HELIX 298 306
 FT TURN 307 307
 FT TURN 311 312
 FT TURN 314 317
 FT HELIX 334 339
 FT TURN 340 341
 FT STRAND 348 353
 FT TURN 354 354
 FT STRAND 355 355
 FT TURN 356 357

```

RESULT 3
LIP5_CANRU STANDARD; PRT; 549 AA.
ID LIP5_CANRU
AC P32949;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Lipase 5 precursor (EC 3.1.1.3).
GN Lip5.
NC Candiida rugosa (Yeast) (Candida cylindracea);
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetaceae; Candida.
OX NCBI_TaxID=5481;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14820;
RC MEDLINE=93178975; PubMed=8440480;
RA Lotti M., Grandori R., Fusetti F., Longhi S., Brocca S.,
RA Tramontano A., Alberghina L.;
RL "Cloning and analysis of Candida cylindracea lipase sequences.";
RT Gene 124:45-55(1993).
RN [2]
RP REVIEW.
RX MEDLINE=96451816; PubMed=9778794;
RA Benjamin S., Pandey A.;
RT "Candida rugosa lipases: molecular biology and versatility in biotechnology.";
RL Yeast 14:1063-1087(1998).
CC -!- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a fatty acid anion.
CC -!- SIMILARITY: Belongs to the type-B carboxylesterase/lipase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X66008; CAA46807.1; -.
DR PIR; JN0553; JN0553.
DR HSP; P32947; LCLE.
DR InterPro; IPR002018; CarboxylaseB.
DR InterPro; IPR000379; Ser_estrs.
DR Pfam; PF00135; Coesterase_1.
DR PROSITE; PS00122; CARBOXYLESTERASE_B_1; 1.
DR PROSITE; PS00941; CARBOXYLESTERASE_B_2; 1.
KW Hydrolase; Lipid degradation; Signal; Glycoprotein; Multigene family.
FT SIGNAL 1 15
FT CHAIN 16 549 LIPASE 5
FT ACT_SITE 224 224 BY SIMILARITY.
FT ACT_SITE 356 356 BY SIMILARITY.
FT ACT_SITE 464 464 BY SIMILARITY.
FT DISULFID 75 112 BY SIMILARITY.
FT DISULFID 283 292 BY SIMILARITY.
FT CARBOHYD 329 329 N-LINKED (GLCNAC.. ) (POTENTIAL) .
FT CARBOHYD 366 366 N-LINKED (GLCNAC.. ) (POTENTIAL) .
SQ SEQUENCE 549 AA; 58420 MW; 0F3D04C9716F6F22 CRC64;
Query Match 84.9%; Score 2431; DB 1; Length 549;
Best local similarity 85.5%; Pred. No. 1.4e-166;
Matches 459; Conservative 32; Mismatches 46; Indels 0; Gaps 0;
Qy 11 LGSVPTAKLANGDTITGNAIINEAFGLPPPEAPPVGNLRPKDPYPYGSLSNGQKFTSYG 70
Db : |||
Db 13 VAAAPTATLANGDTITGLNAIINEAFGLHPPEAPPVGNLRFKDPVPYRGSLSNGQSFTAYG 72
Qy 71 PSCMQQNPEGTFEENLGKTALDLVMQSKVFQAVLPQSEDLTINVRPDPGTAGANLPWM 130
Db 73 PSCMQQNPEGTYEENLPKVALLVMQSKVFQAVLPNSEDLTINVRPDPGTAGANLPWM 132
Qy 131 LWIFGGGFETGSPITFPFAQWTKTSVLIMGKHIIHVAIVNVRSWSGFLAGDDJKAEGSGNA 190

```

```

Db 133 LWIFGGFEIGSPITPPAQWVSKVLMGKPIIHVAVNYRLASFGLAGPDIKAEKSGNA 192
Qy 191 GLKDQRLGMQWADVNIAGFGDPKSVITIFGESAGSMSVLCILWNGDNTYKGLPRAG 250
Db 193 GLKDQRLGMQWADVNIAGFGDPKSVITIFGESAGSMSVLCILWNGDNTYKGLPRAG 252
Qy 251 IMQSGAMVPSDPVDTGTYNEIYDLFVSSAGCGSADKLAELRSASSDTLLDATNNTPGFL 310
Db 253 IMQSGAMVPSDPVDTGTYNEIYDLFVSSAGCGSADKLAELRSASSDTLLDATNNTPGFL 312
Qy 311 AYSSLRSLYPRPDGKNITDMYKLVKRGKVASVPIIGDQNDGDTIFGLSSLVNTTNAQ 370
Db 313 SYTSLSRLYPRPDGANITDMYKLVKRGKVASVPIIGDQNDGDTIFGLSSLVNTTTEAD 372
Qy 371 ARAYFKOSFIHASDAEIDTLMAAYPODITQSGPDTGVLNALTPOFKRISAVLGDIAFIH 430
Db 373 AEAYLRKSFHATDADITALKAAVPSDVTOGSPDTGLNALTPOFKRINAVLGDIAFIH 432
Qy 431 ARBYELNHFQGGTKYSLFSLKSLGSLPIMGTGTHANDIWMQDYLLGSGSVIYNNAFIATD 490
Db 433 SRYFLNHYTGGPKYSLFSLKSLGSLPILGTGTHANDIWMQDYLLGSGSVIYNNAFIATD 492
Qy 491 LDNPNTAGLLVNWPKYTSSSQSGNNLMINALGLYTGKDNFTAGYDALMTNPSSFV 547
Db 493 LDNPNTAGLSVOWPKSTSSQSGNNLMQISALGLYTGKDNFTAGYDALMTNPSSFV 549

```

RESULT 4

```

ID LIP2 CANRU STANDARD; PRT; 549 AA.
AC P32946;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DE Lipase 4 precursor (EC 3.1.1.3).
GN LIP2.
OS Candida rugosa (Yeast) (Candida cylindracea).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5481;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14830;
RX MEDLINE=93178975; PubMed=8440480;
RA Lotti M., Grandori R., Fusetti F., Longhi S., Brocca S.,
RA Tramontano A., Alberghina L.;
RT "Cloning and analysis of Candida cylindracea lipase sequences.";
RL Gene 124:45-55(1993).
RN [2]
RP REVIEW.
RX MEDLINE=98451816; PubMed=9778794;
RA Benjamin S., Pandey A.;
RT "Candida rugosa lipases: molecular biology and versatility in
biotechnology.";
RL Yeast 14:1069-1087(1998).
CC -!- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a
fatty acid anion.
CC -!- SIMILARITY: Belongs to the type-B carboxylesterase/lipase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X66007; CAA46806.1; -
DR PIR; JN0552; JN0552.
DR HSSP; P32947; 1CLE.
DR InterPro; IPR002018; Carboxylesterase.
DR InterPro; IPR000379; Ser_estrs.

```

```

Pfam; PF00135; Coesterase; 1.
DR PROSITE; PS00122; CARBOXYLESTERASE_B_1; 1.
DR PROSITE; PS00941; CARBOXYLESTERASE_B_2; 1.
KW Hydrolase; Lipid degradation; Signal; Glycoprotein; Multigene family.
FT SIGNAL 1 15
FT CHAIN 16 549 LIPASE 4.
FT ACT_SITE 224 224 BY SIMILARITY.
FT ACT_SITE 356 356 BY SIMILARITY.
FT ACT_SITE 464 464 BY SIMILARITY.
FT DISULFID 75 112 BY SIMILARITY.
FT DISULFID 283 292 BY SIMILARITY.
FT CARBOHYD 366 366 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 549 AA; 58570 MW; 39C7160852F7E555 CRC64;

Query Match 82.6%; Score 2366; DB 1; Length 549;
Best Local Similarity 82.5%; Pred. No. 6,2e-162;
Matches 443; Conservative 40; Mismatches 54; Indels 0; Gaps 0;

Qy 11 LGSVPTAKLANGDTITGLNALINEAFILGIPAEPPVGNLRFKDPVYSGSLNGOKFTSYG 70
Db 13 VAAAPTATLANGDTITGLNALINEAFILGIPAEPPVGNLRFKDPVYSGSLNGOKFTSYG 72
Qy 71 PSCMQQNEPTFBEENLGKLTALDLVMSKVFAVLPOSEDCLTINVVRPPGKTGAGANLPVM 130
Db 73 PSCMQMPLGNWDSLLPKAINSIMQSKLFOAVLPNGEDCLTINVVRPPGKTGAGANLPVM 132
Qy 131 LWIFGGFEIGSPITPPAQWVSKVLMGKPIIHVAVNYRLASFGLAGPDIKAEKSGNA 190
Db 133 VWIFGGFEVGGSSLPFAQMITASVLMGKPIIHVSMNYRVSWSGFLAGPDIKAEKSGNA 192
Qy 191 GLKDQRLGMQWADVNIAGFGDPKSVITIFGESAGSMSVLCILWNGDNTYKGLPRAG 250
Db 193 GLHDQRLGMQWADVNIAGFGDPKSVITIFGESAGSMSVLCILWNGDNTYKGLPRAG 252
Qy 251 IMQSGAMVPSDPVDTGTYNEIYDLFVSSAGCGSADKLAELRSASSDTLLDATNNTPGFL 310
Db 253 IMQSGAMVPSDPVDTGTYNEIYDLFVSSAGCGSADKLAELRSASSDTLLDATNNTPGAL 312
Qy 311 AYSSLRSLYPRPDGKNITDMYKLVKRGKVASVPIIGDQNDGDTIFGLSSLVNTTNAQ 370
Db 313 AYPSLRSLYPRPDGDTITDDMFKLVRDKCANVPVPIIGDQNDGDTIFGLSSLVNTTDAQ 372
Qy 371 ARAYFKOSFIHASDAEIDTLMAAYPODITQSGPDTGVLNALTPOFKRISAVLGDIAFIH 430
Db 373 ARQYFKESFIHASDAEIDTLMAAYPSDITQSGPDTGIFNAITPOFKRIAAVLGDIAFIH 432
Qy 431 ARRYFLNHFQGGTKYSLFSLKSLGSLPIMGTGTHANDIWMQDYLLGSGSVIYNNAFIATD 490
Db 433 PRRYFLNHFQGGTKYSLFSLKSLGSLPIMGTGTHANDIWMQDYLLGSGSVIYNNAFIATD 492
Qy 491 LDNPNTAGLLVNWPKYTSSSQSGNNLMINALGLYTGKDNFTAGYDALMTNPSSFV 547
Db 493 LDNPNTAGLLVNWPKYTSSSQSGNNLMINALGLYTGKDNFTAGYDALMTNPSSFV 549

```

RESULT 5

```

ID LIP2 CANRU STANDARD; PRT; 548 AA.
AC P32946;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DE Lipase 2 precursor (EC 3.1.1.3).
GN LIP2.
OS Candida rugosa (Yeast) (Candida cylindracea).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5481;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14830;
RX MEDLINE=92305068; PubMed=1610906;
RA Longhi S., Fusetti F., Grandori R., Lotti M., Vanoni M.,

```


RA Alberghina L.;
RT "Cloning and nucleotide sequences of two lipase genes from *Candida*
RL *cylindracea*.";
RN Biochim. Biophys. Acta 1131:227-232(1992).
RP [2]
RX REVIEW.
RA MEDLINE=98451816; PubMed=9778794;
RT Benjamin S., Pandey A.;
RA "Candida rugosa lipases: molecular biology and versatility in
RT biotechnology.";
RL Yeast 14:1059-1087(1998).
CC -1- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a
CC fatty acid anion.
CC -1- SIMILARITY: Belongs to the type-B carboxylesterase/lipase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X64704; CAA45958.1; -
DR PIR; S32615; S32615.
DR HSP; P32947; ICLE.
DR InterPro; IPR002018; CarbesteraseB.
DR InterPro; IPR000379; Ser esters.
DR Pfam; PF00135; Coesterase; 1.
DR PROSITE; PS00122; CARBOXYLESTERASE B_1; 1.
DR PROSITE; PS00941; CARBOXYLESTERASE B_2; 1.
KW Hydrolyase; Lipid degradation; Signal; Glycoprotein; Multigene family.
FT SIGNAL 1 14
FT CHAIN 15 548
FT ACT_SITE 223 223
FT ACT_SITE 355 355
FT ACT_SITE 463 463
FT DISULFID 74 111
FT DISULFID 282 291
FT CARBOHYD 365 365
FT SEQUENCE 548 AA; 58864 MW; E0DBCF2501E7614 CRC64;
Query Match 80.8%; Score 2314; DB 1; Length 548;
Best Local Similarity 80.8%; Pred. No. 3.3e-158;
Matches 434; Conservative 35; Mismatches 68; Indels 0; Gaps 0;
QY 11 LGSVPTAKLANGDTITGLNAINEAFIPGPPVGNLRFKDPVPSGLNGQKFTSYG 70
Db 12 VAAAPTATLANGDTITGLNAIVNEKFLGIPPAEPVGTLRFPVPPVPSASLNGQKFTSYG 71
QY 71 PSCMQNPGEFTENIGKATLDLVMQSKVQAVLPQSECLTNVVRPPTKAGANLPVM 130
Db 72 PSCMQNPGEFTENIGKATLDLVMQSKVQAVLPQSECLTNVVRPPTKAGANLPVM 131
QY 131 LMTFGGFEIGSTIPPAQMTVSKVLMGKHIIHVAIVNVRVSWGFLAGDDIKAEGSGNA 190
Db 132 LMTFGGFEIGSTIPPAQMTVSKVLMGKHIIHVAIVNVRVSWGFLAGDDIKAEGSGNA 191
QY 191 GLKQDRLGMQWADNADNAGFGGDPFSKVTIFGESAGSMVLCHELLWNGDNTYKGPFLFRAG 250
Db 192 GLHQRQLAMQWADNADNAGFGGDPFSKVTIYESAGSMVLCHELLWNGDNTYKGPFLFRAG 251
QY 251 IMQSGAMVPDVPDVTGNYEIIYDLFVSSACGSGASDKLACLRSSDPTLADNTNPPGFL 310
Db 252 IMQSGAMVPDVPDVTGNYEIIYDLFVSSACGSGASDKLACLRSSDPTLADNTNPPGFL 311
QY 311 AVSLRLSYLPRPDGKNTIDMYKLVDRDGKYSVPVLIIGDNDGDTIFGLSSLNVTNAQ 370
Db 312 AVSLRLSYLPRPDGKNTIDMYKLVDRDGKYSVPVLIIGDNDGDTIFGLSSLNVTNAQ 371
QY 371 ARAYFKQSFHASDAEIDTLMAYPQDITQSGPPTGVNLTQPKRIASVILGDLAFIH 430
Db 372 ARAYFKQSFHASDAEIDTLMAYPQDITQSGPPTGVNLTQPKRIASVILGDLAFIH 431

QY 431 ARRYFLNHFQGTGKYSFLSKQLSCLPIMCTFHANDIWDYLLGSGSVIYNNAFIATFD 490
Db 432 ARRYFLNHFQGTGKYSFLSKQLSCLPIMCTFHANDIWDYLLGSGSVIYNNAFIATFD 491
QY 491 LDPTAGLLVNPVKYTSQSSQGNLMMINAGLYTGKDNFRTAGYDALMTNPSFFV 547
Db 492 LDPNKAGLWNTNPTYTSSQSSQGNLMMINAGLYTGKDNFRTAGYDALMTNPSFFV 548
RESULT 6
LIP1_GEOCN STANDARD; PRT; 563 AA.
ID LIP1_GEOCN
AC P15753;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Lipase 1 precursor (EC 3.1.1.3) (GCL I).
GN LIP1.
OS Geotrichum candidum (Oospora lactis).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Galactomyces.
OX NCBI_TaxID=27317;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=ATCC 34614;
RX MEDLINE=90110016; PubMed=2481674;
RA Shimada Y., Sugihara A., Tominaga Y., Iizumi T., Tsunasawa S.;
RT "cDNA molecular cloning of Geotrichum candidum lipase.";
RL J. Biochem. 106:383-388(1989).
RN [2]
RP SIMILARITY TO CARBOXYLESTERASES.
RC MEDLINE=90329888; PubMed=2115773;
RA Slabas A.R., Windust J., Sidebottom C.M.;
RT "Does sequence similarity of human choline esterase, Torpedo
RT acetylcholine esterase and Geotrichum candidum lipase reveal the
RL active site serine residue?"
RN Biochem. J. 269:279-280(1990).
RN [3]
RP CHARACTERIZATION.
RC MEDLINE=9525295; PubMed=7737187;
RA Bertolini M.C., Schrag J.D., Cygler M., Ziomek E., Thomas D.V.,
RA Vernet T.;
RT "Expression and characterization of Geotrichum candidum lipase I gene.
RT Comparison of specificity profile with lipase II.";
RL Eur. J. Biochem. 228:863-869(1995).
CC -1- FUNCTION: Hydrolyzes all ester bonds in triglyceride and displays
CC a high affinity for triolein. For unsaturated substrates having
CC long fatty acyl chains (C18:2 cis-9, cis-12 and C18:3 cis-9,
CC cis-12, cis-15) GCL I shows higher specific activity than GCL II,
CC whereas GCL II shows higher specific activity against saturated
CC substrates having short fatty acid chains (C8, C10, C12 and C14).
CC -1- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a
CC fatty acid anion.
CC -1- SUBUNIT: Monomer.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to the type-B carboxylesterase/lipase family.
DR PIR; P0492; ACUGC.
DR InterPro; IPR002018; CarbesteraseB.
DR InterPro; IPR000379; Ser esters.
DR Pfam; PF00135; Coesterase; 1.
DR PROSITE; PS00122; CARBOXYLESTERASE B_1; 1.
DR PROSITE; PS00941; CARBOXYLESTERASE B_2; 1.
KW Hydrolyase; Lipid degradation; Glycoprotein; Signal;
FT SIGNAL 1 19
FT CHAIN 20 563
FT MOD_RES 20 20
FT ACT_SITE 236 236
FT ACT_SITE 373 373
FT ACT_SITE 482 482
FT DISULFID 80 124
FT DISULFID 295 307
FT PYRROLIDONE CARBOXYLIC ACID.
FT LIPASE 1.

Fri Aug 6 10:53:13 2004

us-09-943-857-4.rsp

```

Query Match          37.6%; Score 1078; DB 1; Length 563;
Best Local Similarity 41.6%; Pred. No. 1.1e-59;
Matches 233; Conservative 83; Mismatches 202; Indels 42; Gaps 8;

QY 8 AGRGVSPTAKLANGDTITGLNALINEAFGLGPFPAEPVGNLRFKDPVPYSGSLNGQKFT 67
DB 15 AGVLAQAPPSLNGNEVIGSVLEKGVDTFKGPFADPPLNDLRFKHPQPFQFTGSGYQGLKAN 74

QY 68 SYGSPCMQNPETFEENLGKTAALDLMV-OSKVF-----QAVLPQSEDCLTIN 114
DB 75 DFPACQMLDPGNSL-----TLDDKALGLAKVPIPEFRGLPYDMAKVTSMNEDCLYLIN 128

QY 115 VVRPPGKAGANLPMVLWTFGGGFEIGSPITRPPAQMTWTKSVLMGKHIIHVAVNVYVASW 174
DB 129 VFRPAGTKPKDALPVMWVIYGGAPVTYSSAAVPGNSVYKESINQGPVVFVSINRYTGP 188

QY 175 GFLAGDDIIRKAGSGNAGLKDQRLGMQVADNIAGFGDGPSPKVTIPGESAGMSVLCILHIW 234
DB 189 GFLGGDAITRAGNTNAGLHDQKGLWVSDNIANFGGDPDKVMIFGESAGMSVAHOLIA 248

QY 235 NGDNTYKGLPFRAGIMQSGAMVPSDPDGTGVGNEYIDLFSVAGC---GSASDKLACL 291
DB 249 YGDDNTYNGKKLFHSAILOSGLPLPYHDSVSGVDISYNRFAQYAGCDTSASANDTLECL 308

QY 292 RSASSDTLLDATNN-----TPGFLAYSSLRSLSVLPDPDGKNIITDDMYKLVRDGKYA 342
DB 309 RSKSSSVLHDAQNSYDLKDLFGLLPQFLFG-----PRPDGNIIPDAAYELFRSGRYA 361

QY 343 SVPIIGDQNDGDTIFGLSSINVTNAQARAYFKOSFIHSDADRIDTMAAYPODITQGS 402
DB 362 KVPVISQNEDEGTAFAFAPVALNATTPHVKKWLQYIFYDASEASIDRVLSLPQTLSVGS 421

QY 403 PFDGVLNALTPOPKRISAVLGDIAFHARRYFLNHPGGTKYSFLSKQLSGL-PIMGTF 461
DB 422 PFRGILNALTPOPKRVAAILSDMLFQSPKRWMLSATKDVNRWYLSLTHLHNLPFFLGTF 481

QY 462 HANDIVQDYLGGSSVLYNNAFAFAFDLPDNTAGLLVNNPKYTSSTSSGNNLMINAL 521
DB 482 HGNELIFQFNVINIGPANSYLYRFLISFANHHPDNPVGNLLQWDQYT---DECKEMLEIHT 538

QY 522 GLYTGKCNFRTAGYDALMTN 541
DB 539 DNWMTDDYRIEGISNPFET 558

RESULT 8
ACCS TORCA          STANDARD;          PRT;          586 AA.
AC P04058;
DT 01-NOV-1986 (Rel. 03, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Acetylcholinesterase precursor (EC 3.1.1.7) (AChE).
OS Torpedo californica (Pacific electric ray).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Squala; Hynostomales; Pristiorajae; Batoidae;
OC Torpediniformes; Torpedinoidei; Torpedinidae; Torpedo.
OX NCBI_TaxID=7787;
[1]
RN SEQUENCE OF 10-586 FROM N.A.
RX MEDLINE=86118676; PubMed=3753747;
RA Schumacher M., Camp S., Maulet Y., Newton M., McPhee-Quigley K.,
RA Taylor S.S., Friedmann T., Taylor P.;
RT "Primary structure of Torpedo californica acetylcholinesterase
RT deduced from its cDNA sequence.";
RL Nature 319:407-409(1986).
[2]
RN SEQUENCE OF 1-9 FROM N.A.
RX MEDLINE=89066695; PubMed=3198606;
RA Schumacher M.;
RT "Multiple messenger RNA species give rise to the structural diversity
RT in acetylcholinesterase.";
RL J. Biol. Chem. 263:18979-18987(1988).

```

```

RN [3]
RX SEQUENCE OF 552-558.
RX MEDLINE=88087239; PubMed=3335534;
RA Gibney G., MacPhee-Quigley K., Thompson B., Vedvick T., Low M.G.,
RA Taylor S.S., Taylor P.;
RT "Divergence in primary structure between the molecular forms of
RT acetylcholinesterase.";
RL J. Biol. Chem. 263:1140-1145(1988).
[4]
RN ALTERNATIVE SPLICING.
RX MEDLINE=90166618; PubMed=2306366;
RA Maulet Y., Camp S., Gibney G., Rachinsky T.L., Ekstroem T.J.,
RA Taylor P.;
RT "Single gene encodes glycopospholipid-anchored and asymmetric
RT acetylcholinesterase forms: alternative coding exons contain inverted
RT repeat sequences.";
RL Neuron 4:289-301(1990).
[5]
RN DISULFIDE BONDS.
RX MEDLINE=87008586; PubMed=3759980;
RA McPhee-Quigley K., Vedvick T.S., Taylor P., Taylor S.S.;
RT "Profile of the disulfide bonds in acetylcholinesterase.";
RL J. Biol. Chem. 261:13565-13570(1986).
[6]
RN STRUCTURE OF THE GPI-ANCHOR.
RX MEDLINE=94079692; PubMed=8257440;
RA Mehler A., Varon L., Silman I., Homans S.W., Ferguson M.A.;
RT "Structure of the glycosyl-phosphatidylinositol membrane anchor of
RT acetylcholinesterase from the electric organ of the electric-fish,
RT Torpedo californica.";
RL Biochem. J. 296:473-479(1993).
[7]
RN GPI-ANCHOR.
RX MEDLINE=96176849; PubMed=8597567;
RA Bucht G., Hjalmarsson K.;
RT "Residues in Torpedo californica acetylcholinesterase necessary for
RT processing to a glycosyl phosphatidylinositol-anchored form.";
RL Biochim. Biophys. Acta 1292:223-232(1996).
[8]
RN MUTAGENESIS.
RX MEDLINE=91017542; PubMed=2217185;
RA Gibney G., Camp S., Dionne M., McPhee-Quigley K., Taylor P.;
RT "Mutagenesis of essential functional residues in
RT acetylcholinesterase.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:7546-7550(1990).
[9]
RX X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RX MEDLINE=91343928; PubMed=1678899;
RA Sussman J.L., Harel M., Frolow F., Oefner C., Goldman A., Tokar L.,
RA Silman I.;
RT "Atomic structure of acetylcholinesterase from Torpedo californica: a
RT prototypic acetylcholine-binding protein.";
RL Science 253:872-879(1991).
[10]
RX X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS) OF COMPLEX WITH FASCICULIN.
RX MEDLINE=96363673; PubMed=8747462;
RA Harel M., Kleywegt G.J., Ravelli R.B., Silman I., Sussman J.L.;
RT "Crystal structure of an acetylcholinesterase-fasciculin complex:
RT interaction of a three-fingered toxin from snake venom with its
RT target.";
RL Structure 3:1355-1366(1995).
[11]
RX X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=97143314; PubMed=8989325;
RA Raves M.L., Harel M., Pang Y.P., Silman I., Kozikowski A.P.,
RA Sussman J.L.;
RT "Structure of acetylcholinesterase complexed with the nootropic
RT alkaloid, (-)-huperzine A.";
RL Nat. Struct. Biol. 4:57-63(1997).
[12]
RX X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
RX MEDLINE=99249780; PubMed=10231521;
RA Bartolucci C., Perola E., Cellai L., Brufani M., Lamba D.;

```


Fri Aug 6 10:53:13 2004

AC 042275;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Acetylcholinesterase precursor (EC 3.1.1.7) (ACHE).
 OS Electrophorus electricus (Electric eel).
 DE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Gymnotiformes;
 OC Electrophoridae; Electrophorus.
 OX NCBI_TaxID=8005;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98070504; PubMed=9407087;
 RA Simon S., Massoulie J.;
 RT Cloning and expression of acetylcholinesterase from Electrophorus.
 RT Splicing pattern of the 3' exons in vivo and in transfected mammalian
 RT cells.";
 RL J. Biol. Chem. 272:33045-33055(1997).
 CC -1- FUNCTION: Rapidly hydrolyzes choline released into the synapse.
 CC -1- CATALYTIC ACTIVITY: Acetylcholine + H₂O = choline + acetate.
 CC -1- SUBUNIT: DIMERS AND COLLAGEN-TAILED FORMS, IN WHICH CATALYTIC
 CC -1- TETRAMERS ARE ASSOCIATED WITH ANCHORING PROTEINS THAT ATTACH THEM
 CC TO THE BASAL LAMINA OR TO CELL MEMBRANES. IN THE COLLAGEN-TAILED
 CC FORMS, SUBUNITS ARE ASSOCIATED WITH A SPECIFIC COLLAGEN, COLQ,
 CC WHICH TRIGGERS THE FORMATION OF ISOFORM T TETRAMERS FROM DIMERS.
 CC -1- MISCELLANEOUS: NO OTHER ISOFORMS EXIST. THIS PROTEIN CORRESPONDS
 CC TO THE T ISOFORM IN OTHER SPECIES.
 CC -1- SIMILARITY: Belongs to the type-B carboxylesterase/lipase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF030422; AAB86606.1; .
 DR HSP; P04058; ISOM
 DR InterPro: IPR002018; CarbestraseB.
 DR InterPro: IPR000997; Cholinesterase.
 DR InterPro: IPR000379; Ser. estera.
 DR Pfam: PF00135; Coesterase; 1.
 DR PRINTS: PR00878; CHOLNESTRASE.
 DR PROSITE: PS00122; CARBOXYLESTERASE_B_1; 1.
 DR PROSITE: PS00941; CARBOXYLESTERASE_B_2; 1.
 DR Hydrolase; Serine esterase; Synapse; Membrane; Nerve; Muscle; Signal;
 KW Neurotransmitter degradation; Glycoprotein.
 FT SIGNAL 1 23
 FT CHAIN 24 633 ACETYLCHOLINESTERASE.
 FT ACT_SITE 225 225 BY SIMILARITY.
 FT ACT_SITE 352 352 BY SIMILARITY.
 FT ACT_SITE 494 494 BY SIMILARITY.
 FT DISULFID 91 118 BY SIMILARITY.
 FT DISULFID 279 290 BY SIMILARITY.
 FT DISULFID 427 579 BY SIMILARITY.
 FT DISULFID 630 630 INTERCHAIN (BY SIMILARITY).
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 283 283 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 368 368 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 511 511 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 591 591 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 633 AA; 71814 MW; FC92FE7E4ADB84C3 CRC64;
 Query Match 17.1%; Score 491; DB 1; Length 633;
 Best Local Similarity 30.3%; Pred. No. 1.5e-27;
 Matches 157; Conservative 84; Mismatches 192; Indels 86; Gaps 22;
 QY 10 RLGSVPTAKLANGTITGLNAINEARLIGIPFAEPVGNLRFKDP---VPYSGSLNGQKF 66
 DB 34 RLGSVPTAKLANGTITGLNAINEARLIGIPFAEPVGNLRFKDP---VPYSGSLNGQKF 66
 DB 34 RLGSVPTAKLANGTITGLNAINEARLIGIPFAEPVGNLRFKDP---VPYSGSLNGQKF 66

QY 67 TSYGSCMOQNEPTFEENLTKALDLYMQSKVQAVLPQSEBCLTINVRPPTKAGAN 126
 DB 88 -----PSACYQYDYDTYTPGFSG-----TEMNPNRMMSDEDCLYLN-VWVPATPRPHN 133
 QY 127 LPVLMWIFGGGPEIGSPITFPDPAQWVTKSVLMKGHIH-----VAVNYRVASWAGLAGD 180
 DB 134 LTVMWIYGGGYSSSSSL-----DVIDGRYLHSEKVVVWSMRYRVSFAFGELALN 184
 QY 181 DIKASGSGNAGLKQRLGMQWADNIAGRGDPSKVTIFGESAGSMVLCILWINDGDMT 240
 DB 185 G-SAEAPGNVGLDQRLALQWVQDNIHFPGNPKQVTIFGESAGASVGVGHLLSPD---- 239
 QY 241 YGKPLFRAGIMQSGMWPESDPVDCITYGNEIYDLFVS-----SAGC--GSASDKLACLSAS 295
 DB 240 --SRPKFTRAILQSG--VPNGPWRVTVSFEARRRAIKRLGVCCPDGNDTDLDCLSKQ 295
 QY 296 SDTLIDATNTPGFLAYSSL-RLSYLPRDPGKNITDDMYKLVRDQGYASVPVVIIGQND 354
 DB 296 PQDLIDQEWLV---LPFSGLFREFVPVVDGVVFPDTPPEAMLSGNFKDTQILLGVNQNE 352
 QY 355 GTIF-----GLSSLN--VTTNAQARAYKQSPFIHASDAEIDTLMAAYPDITQGSFFDT 406
 DB 353 GSYFLIYGAPGFSKDNESLITREDFLQGVKMSVPHANEIGLEAVILQYTDWMDENPIKN 412
 QY 407 GVNLALTPOPKRISAVLGDLAFHARRYFLNHPQGGTKYSFLSKQLSGLPIMGTFRANDI 466
 DB 413 R-----EAMDDIVGDHNVVCPLOHFAKMY---AQVSILOQ-----TGTASQGNL 454
 QY 467 VMQDYLLGSGSYNNNAFAFATDLDENTAGLVNMPKY 505
 DB 455 GWGN--SGSASNSGNSQSVLYWFDHRASNLV--WPEW 489
 RESULT 10
 ACES_TORMA STANDARD; PRT; 590 AA.
 ID ACES_TORMA STANDARD; PRT; 590 AA.
 AC P07692;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Acetylcholinesterase precursor (EC 3.1.1.7) (ACHE).
 OS Torpedo marmorata (Marbled electric ray).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
 OC Elasmobranchii; Squalia; Hymnosqualia; Priestiorajae; Batoida;
 OC Torpediniformes; Torpedinoidei; Torpedinidae; Torpedo.
 OX NCBI_TaxID=7788;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Electric organ;
 RX MEDLINE=88004392; PubMed=2820709;
 RA Sikorav J.-L., Krejci E., Massoulie J.;
 RT "cDNA sequences of Torpedo marmorata acetylcholinesterase: primary
 RT structure of the precursor of a catalytic subunit; existence of
 RT multiple 5'-untranslated regions.";
 RL EMBO J. 6:1865-1873(1987).
 RN [2]
 RP ALTERNATIVE SPLICING.
 RX TISSUE=Electric organ;
 RX MEDLINE=89030590; PubMed=3181125;
 RA Sikorav J.-L., Duval N., Anselmet A., Bon S., Krejci E., Legay C.,
 RA Osterlund M., Reimund B., Massoulie J.;
 RT "Complex alternative splicing of acetylcholinesterase transcripts in
 RT Torpedo electric organ; primary structure of the precursor of the
 RT glycolipid-anchored dimeric form.";
 RL EMBO J. 7:2983-2993(1988).
 RN [3]
 RP SUBUNITS INTERACTIONS.
 RX TISSUE=Electric organ;
 RX MEDLINE=92371432; PubMed=1380451;
 RA Duval N., Krejci E., Grassi J., Coussen F., Massoulie J., Bon S.;
 RT "Molecular architecture of acetylcholinesterase collagen-tailed
 RT forms; construction of a glycolipid-tailed tetramer.";
 RL EMBO J. 11:3255-3261(1992).


```
CC carboxylic acid anion.
CC -i- SUBUNIT: Homotetramer. The tetramer is composed of two dimers. The
CC two subunits in a dimer are linked by a disulfide bond.
CC -i- TISSUE SPECIFICITY: Present in most cells except erythrocytes.
CC -i- MISCELLANEOUS: Cholinesterase is highly reactive with
CC organophosphate esters.
CC -i- SIMILARITY: Belongs to the type-B carboxylesterase/lipase family.
CC HSP; P21836; IMAA.
DR InterPro; IPR002018; Carboxylesterase.
DR InterPro; IPR000997; Cholinesterase.
DR Pfam; PF00135; Coesterase; 1.
DR PRINTS; PR00878; CHOLINESTRASE.
DR PROSITE; PS00122; CARBOXYLESTERASE_B_1; 1.
DR PROSITE; PS00941; CARBOXYLESTERASE_B_2; 1.
KW Hydrolase; Serine esterase; Glycoprotein.
FT ACT SITE 198
FT ACT SITE 325 BY SIMILARITY.
FT ACT SITE 438 BY SIMILARITY.
FT DISULFID 65 92 BY SIMILARITY.
FT DISULFID 252 263 BY SIMILARITY.
FT DISULFID 400 519 BY SIMILARITY.
FT DISULFID 571 571 INTERCHAIN (BY SIMILARITY).
FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .).
FT CARBOHYD 106 106 N-LINKED (GLCNAC. . .).
FT CARBOHYD 241 241 N-LINKED (GLCNAC. . .).
FT CARBOHYD 256 256 N-LINKED (GLCNAC. . .).
FT CARBOHYD 341 341 N-LINKED (GLCNAC. . .).
FT CARBOHYD 455 455 N-LINKED (GLCNAC. . .).
FT CARBOHYD 481 481 N-LINKED (GLCNAC. . .).
FT CARBOHYD 486 486 N-LINKED (GLCNAC. . .).
SQ SEQUENCE 574 AA; 65641 MW; 07755BE9FB9CB33E CRC64;

Query Match 16.8%; Score 482; DB 1; Length 574;
Best Local Similarity 30.4%; Pred. No. 5.6e-27;
Matches 162; Conservative 78; Mismatches 193; Indels 100; Gaps 23;

QY 22 GDTITGLNAINEAFGLGIPFAEPVGNLRFDPVYSGSLNGQKFTSYGSCMQQNPRGT 81
DB 22 GGTVT-----AFLGIPYAPPLGRFRFKFQSLTWSNIWATKYANSC-YQNTQS 72

QY 82 FEENLGTALDVMQSKVQAVLPQSEDLTINVRP-PGTGAGANLPLVMIWIFGGGPEI 140
DB 73 PPGFLG-----SEMNPNTELSEDLVNLVWIPAPKPK---NATVMIWIYGGGQPT 120

QY 141 GSPTIFFPAQMYKSVLMGK-----HIIHVAIVNVASWGLAGDDIKAGSGNAGLKD 194
DB 121 GTSSLJ-----PVYDGKFLARVERVIVSMNVRVGALGLALSE-NPEAPGNMGLFD 170

QY 195 QRLGMQWADVADIAFGDPSKVTIFGESAGNSVLCHLWNGDNTYKGLKPLFRAGIMQS 254
DB 171 QQLALQWQKNIARFGGNPRSVTLFGESAGASVSLHL-----SPRSQPLFTRAILQS 224

QY 255 GAVVPSDPVDTYGNIEYDLFVSS-AGCG--SASDKLACLRSASDITLDTNNTPGFLA 311
DB 225 GSSNAPWAVTSIYEARNRTLTLAKMGCSRDNETEMIKCLRDKDPQEL---LNEVFVVP 281

QY 312 YSS-LRLSVLPRPCKNITDDMYKLVRGKYASVPIIGDQNDCTIF-----GLSLIN 364
DB 282 YDTLLSVNFGPTVDGDFITLQGLQFRTQLLVGVNKGDEGTAFLVYGAPGFSKDN 341

QY 365 --VITNAQARAYFKQSFHASDAEIDTLMAAY-----FQDITQGSFPFDTGVLNALT 414
DB 342 NSIITRKEFQEGLKIFFRVSFEGRESILFHYMDLWDQRAENYREALDDVVDYNIICP 401

QY 415 --QFKRISAVIGDLAFIARRIFLNFHQGTYKSFQSLGSLPMTGFHANDIYVQDYL 472
DB 402 ALEFTRKFSGLNDAPF---YFHEI-----RSTKLPPWPEWGMVGMHGYEI---BFV 445

QY 473 LG---SGSVIYNNA-----FIAPATLDEN-TAGLLVNNPKVTSQQ 510
DB 446 FGLPLERRVNYTRAEIILSRIMKRWANPAKYNFNQNTQNNSTRNPFVKSTEQ 498
```

```
RESULT 12
ACCS MOUSE
ID _ACCS_MOUSE STANDARD; PRT; 614 AA.
AC P21836;1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE "Acetylcholinesterase precursor (BC 3.1.1.7) (AChE)".
GN ACHE
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
MEDLINE=90380429; PubMed=2400605;
RA Rachinsky T.B., Camp S., Li Y., Ekstrom T.J., Newton M., Taylor P.;
RT "Molecular cloning of mouse acetylcholinesterase: tissue distribution
of alternatively spliced mRNA species.";
RL Neuron 5:317-327(1990).
RN [2]
RP SEQUENCE FROM N.A.
STRAIN=129/SV;
MEDLINE=21138439; PubMed=11239002;
RA Wilson M.D., Riemer C., Martindale D.W., Schnupf P., Boright A.P.,
RA Cheung T.L., Hardy D.M., Schwartz S., Scherer S.W., Tsui L.-C.,
RA Miller W., Koop B.F.;
RT "Comparative analysis of the gene-dense ACHE/TFP2 region on human
chromosome 7q22 with the orthologous region on mouse chromosome 5.";
RL Nucleic Acids Res. 29:1352-1365(2001).
RN [3]
RP SEQUENCE FROM N.A.
TISSUE=Retina;
MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
RA Datchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Raha S.S., Lochlanno N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Maman A., Rodriguez S., Sanchez A.,
RA Whiting M., Maman A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP INTERACTION WITH PRIMA1.
MEDLINE=21664287; PubMed=11804574;
RA Perrier A.L., Maesoulie J., Krejci E.;
RT "PRIMA: the membrane anchor of acetylcholinesterase in the brain.";
RL Neuron 33:275-285(2002).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (3.2 ANGSTROMS) OF COMPLEX WITH FASCICULIN.
MEDLINE=96067648; PubMed=8521480;
RA Bourne Y., Taylor P., Marchot P.;
RT "Acetylcholinesterase inhibition by fasciculin: crystal structure of
the complex.";
RL Cell 83:503-512(1995).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).
MEDLINE=99115643; PubMed=9915834;
RA Bourne Y., Taylor P., Bougis P.E., Marchot P.;
RT "Crystal structure of mouse acetylcholinesterase. A peripheral site-
```


occluding loop in a tetrameric assembly.";

RT J. Biol. Chem. 274:2963-2970(1999).

CC -!- FUNCTION: Rapidly hydrolyzes choline released into the synapse.

CC -!- CATALYTIC ACTIVITY: Acetylcholine + H(2)O = choline + acetate.

CC -!- SUBUNIT: Isoform H generates GPI-anchored dimers; disulfide

CC linked. Isoform T generates multiple structures, ranging from

CC monomers and dimers to collagen-tailed and hydrophobic-tailed

CC forms, in which catalytic tetramers are associated with anchoring

CC proteins that attach them to the basal lamina or to cell

CC membranes. In the collagen-tailed forms, isoform T subunits are

CC associated with a specific collagen, COLQ, which triggers the

CC formation of isoform T tetramers, from monomers and dimers (By

CC similarity). Interacts with PRIMA1. The interaction with PRIMA1 is

CC required to anchor it to the basal lamina of cells and organize

CC into tetramers.

CC -!- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=2;

CC Name=T;

CC IsoId=P21836-1; Sequences=Displayed;

CC Name=H;

CC IsoId=P21836-2; Sequence=Not described;

CC Note=No experimental confirmation available;

CC -!- TISSUE SPECIFICITY: Predominates in most expressing tissues except

CC erythrocytes where a glycopospholipid-attached form of AChE

CC predominates.

CC -!- MISCELLANEOUS: Synapses usually contain asymmetric molecules of

CC cholinesterase, with a collagen-like part disulfide-bonded to the

CC catalytic part. A different, globular type of cholinesterase

CC occurs on the outer surfaces of cell membranes, including those of

CC erythrocytes.

CC -!- MISCELLANEOUS: This is the catalytic subunit of an asymmetric or

CC soluble form of AChE.

CC -!- SIMILARITY: Belongs to the type-B carboxylesterase/lipase family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X56518; CAA39867.1; -;

DR EMBL; AF312033; RAK28816.1; -;

DR EMBL; BC046327; AAH46327.1; -;

DR PIR; JH0314; JH0314

DR PDB; 1MAH; 03-APR-96.

DR PDB; 1MAA; 20-APR-99.

DR PDB; 1C2B; 29-DEC-99.

DR PDB; 1C2O; 19-JAN-00.

DR PDB; 1J06; 04-FEB-03.

DR PDB; 1J07; 04-FEB-03.

DR PDB; 1NSM; 04-FEB-03.

DR PDB; 1NSR; 04-FEB-03.

DR MGD; MGI:87876; Ache.

DR GO; GO:0045202; C:synaptic junction; IDA.

DR InterPro; IPR002018; CarboxylesteraseB.

DR InterPro; IPR000997; Cholinesterase.

DR Pfam; PF00135; Coesterase; 1.

DR PRINTS; PR00878; CHOLINESTRASE.

DR PROSITE; PS00122; CARBOXYLESTERASE B_1; 1.

DR PROSITE; PS00941; CARBOXYLESTERASE B_2; 1.

KW Hydrolase; Serine esterase; Synapse; Membrane; Nerve; Muscle; Signal;

KW Neurotransmitter degradation; Glycoprotein; Alternative splicing;

KW 3D-structure.

FT SIGNAL 1 31

FT CHAIN 32 614 ACETYLCHOLINESTERASE.

FT ACT_SITE 234 234

FT ACT_SITE 365 365

FT ACT_SITE 478 478

FT DISULFID 100 127

FT DISULFID 288 303

560 FT DISULFID 440

611 FT DISULFID 611

296 FT CARBOHYD 296

381 FT CARBOHYD 381

495 FT CARBOHYD 495

39 FT HELIX 37

43 FT STRAND 40

44 FT TURN 44

45 FT STRAND 46

49 FT STRAND 51

55 FT TURN 56

57 FT STRAND 58

67 FT STRAND 69

69 FT HELIX 74

76 FT TURN 77

78 FT STRAND 83

83 FT STRAND 90

92 FT STRAND 94

94 FT STRAND 99

100 FT TURN 109

110 FT TURN 112

115 FT HELIX 116

117 FT STRAND 123

124 FT STRAND 129

135 FT STRAND 143

149 FT TURN 153

155 FT TURN 159

160 FT HELIX 162

164 FT HELIX 166

173 FT STRAND 176

180 FT HELIX 185

189 FT TURN 193

194 FT HELIX 202

217 FT HELIX 218

221 FT TURN 222

222 FT STRAND 223

233 FT TURN 234

234 FT HELIX 235

244 FT TURN 247

250 FT TURN 251

252 FT STRAND 255

259 FT STRAND 270

270 FT HELIX 272

285 FT TURN 286

287 FT HELIX 297

306 FT HELIX 309

315 FT HELIX 316

319 FT STRAND 333

333 FT TURN 353

354 FT STRAND 356

362 FT TURN 363

364 FT HELIX 367

370 FT TURN 371

373 FT TURN 375

376 FT HELIX 387

397 FT TURN 399

400

Query Match 16.8%; Score 481; DB 1; Length 614;

Best Local Similarity 30.2%; Pred. No. 7.3e-27;

Matches 162; Conservative 78; Mismatches 177; Indels 120; Gaps 26;

QY 35 AFLGIPFAEPVGNLRFKDPV---PVSGLNQKFTSYGFCMQ-----QNPEGTFEEN 85

DB 62 AFLGIPFAEPVGNLRFKDPV---PVSGLNQKFTSYGFCMQ-----QNPEGTFEEN 118

QY 86 LGKTAIDLVMQSKVFQAVLPQSEDCLTINVRPPGKTGAGANLPVMLWIFGGGFEGSPPT 145

DB 119 PNREL-----SEDCLYLNVTWTFYPRPASPT-PVLIWIYGGGFYSGAASL 161

QY 146 FPPAQWTKSVLMGKHI-----IHVANNYRVASNGFLAGDDIKAEAGSGNAGLKQRLGM 199

DB 162 -----DVEDGRFLAQVEGAVLVSNNRYGVTFGLALPGSR-EAPGNVGLLDQRLAL 211

QY 200 QWADNIAGGDPKVTIFGESAGSMVLCILWINDGNTYKGLPFRAGIMOSGAMVP 259
DB 212 QWQENTAAAGGDDPMSVTLFESAGSAASVGMHIL-----SLPSRSLFHRVLOSG--TP 263
QY 260 SDP---VDGTGNEIYDLFVSSAGC-----GSASDKLACLRSSASTLTDATNTTGPFL 310
DB 264 NGPWATVSAGARRATLLARLVGCPGAGGNDTELIACLRTPAQDLVDHEWHV--L 320
QY 311 AYSSL-RLSYLPRDGNITDDMTKLYRDGKYASVPVIGDNDGCTIF-----GLSSL 363
DB 321 PQSIFRFSFVVDGDFSLDTPBALINTGDFQDLQVLGVVXDEGSYFLVYGVPGFSKD 380
QY 364 N--VTTNAQARAYFKQGFIIASDAEIDTLMAAY-----PDITQGSFPDFTGLV---NALT 413
DB 381 NESLISRAQFLAGVIRIGVQASDLAAENVLHYDMLHPEDPHLRDAMSVAVDGHNVC 440
QY 414 PQKRISAVIGDLAFIHHARRYFLNHFGQGTYSFLSKOLSGL-----PI-MGFTHANDIVWQ 469
DB 441 P---VAQLAGRLA-----AQAARVAYIFEHRASTLTWPLMWGVPHGYEI--- 482
QY 470 DYLLG---SGSVIYNNAFIATDL-----DNTA--GLLVNWPKYTSSSQ 510
DB 483 EFIFGLDPSLNTTTEIRIPAQRIMKYTNWTFARTGDPNDPRDSKSPQMPPTTAAQ 539

RESULT 13

CHLE_MOUSE
ID CHLE_MOUSE STANDARD; PRT; 603 AA.
AC Q03311;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Cholinesterase precursor (EC 3.1.1.8) (Acylcholine acylhydrolase)
DE (Choline esterase II) (Butyrylcholine esterase)
DE (Pseudochoolinesterase).
GN BCHE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90380429; PubMed=240605;
RA Rachinsky T.L., Camp S., Li Y., Ekstrom T.J., Newton M., Taylor P.;
RT "Molecular cloning of mouse acetylcholinesterase: tissue distribution
of alternatively spliced mRNA species.";
RL Neuron 5:317-327(1990).
RN [2]
RP SEQUENCE OF 97-237 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=91201348; PubMed=2016308;
RA Artigas M., Chatonnet A., Masson P., Newton M., Vaughan T.A.,
RA Bartels C.F., Nogueira C.P., la Du B.N., Lockridge O.;
RT "Use of the polymerase chain reaction for homology probing of
butyrylcholinesterase from several vertebrates.";
RL J. Biol. Chem. 266:6966-6974(1991).
CC -1- CATALYTIC ACTIVITY: An acylcholine + H(2)O = choline + a
CC carboxylic acid anion.
CC -1- SUBUNIT: Homotetramer. The tetramer is composed of two dimers. The
CC two subunits in a dimer are linked by a disulfide bond.
CC -1- TISSUE SPECIFICITY: Present in most cells except erythrocytes.
CC -1- MISCELLANEOUS: Cholinesterase is highly reactive with
CC organophosphate esters.
CC -1- SIMILARITY: Belongs to the type-B carboxylesterase/lipase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; M99492; AAA37328.1; --
DR PIR; S70849; S70849.
DR HSP; P21836; IMAH.
DR MGD; MGI:894278; Bche.
DR DR InterPro; IPR002018; CarboxylesteraseB.
DR DR InterPro; IPR000977; Cholinesterase.
DR DR InterPro; IPR000379; Ser esters.
DR Pfam; PF001135; Coesterase; 1.
DR PRINTS; PR00878; CHOLINESTERASE.
DR PROSITE; PS00122; CARBOXYLESTERASE_B_1; 1.
DR PROSITE; PS00941; CARBOXYLESTERASE_B_2; 1.
KW Hydrolase; Serine esterase; Glycoprotein; Signal.
FT SIGNAL 1 29
FT CHAIN 30 603 CHOLINESTERASE.
FT ACT_SITE 227 227 BY SIMILARITY.
FT ACT_SITE 354 354 BY SIMILARITY.
FT ACT_SITE 467 467 BY SIMILARITY.
FT DISULFID 94 121 BY SIMILARITY.
FT DISULFID 281 292 BY SIMILARITY.
FT DISULFID 429 548 BY SIMILARITY.
FT DISULFID 600 600 INTERCHAIN (BY SIMILARITY).
FT CARBOHYD 86 86 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 135 135 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 270 270 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 370 370 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 484 484 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 510 510 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 129 129 R -> P (IN REF. 2).
SQ SEQUENCE 603 AA; 68521 MW; 719B1B220D1E5367 CRC64;

Query Match 16.6%; Score 476; DB 1; Length 603;
Best Local Similarity 30.7%; Pred. No. 1.6e-26;
Matches 164; Conservative 75; Mismatches 193; Indels 102; Gaps 24;

QY 22 GDTITGLNAINAEFLGIPFAEPVGNLRPKDPVYSGSLNGKQFTSYGSCMQONPEGT 81
DB 51 GGTVT-----AFLGIPYAAQPLGSLRFKPKQPLNKNWPDHINATQYNSC-YQNDQA 101
QY 82 FEENLGTALDLVMQSFQAVLPQSEDCLTINV-VRPPGTKAGANLPVMLWIRGGGFEI 140
DB 102 FPGFQG-----SEMMNPNTNLSDECLYNWIRVPKPK---NATVWVWIIYGGGFQT 149
QY 141 GSPTIPPPAQMVKSVLMGK-----HIHVAVNRYVASWGFAGDDIKAGSGNAGLKD 194
DB 150 GTSSL-----PVYDGKFLARVERVIVVMNRYVGALGFLAFPG-NPDAPGNMGLFD 199
QY 195 QRLGMQWADNIAGFGGDPKVTIFGESAGSMVLCILWINDGNTYKGLPFRAGIMOS 254
DB 200 QQLALQWQVQRIAAFGGNPKSITIFGESAGAAVSLSHL-----CPQSYPLFTRAI 253
QY 255 GAMVPSDPVDGTGNEIYDLFVSSA---GCGSAS--DKLACLRSSASTLTDATNTTGF 309
DB 254 GS--SNAPWAVKPEEARNTLTIAKFTGSKENEMIKCLRSKDQEIIL--RNERFVL 309
QY 310 LAYSSLRLSYLPRDGNITDDMTKLYRDGKYASVPVIGDNDGCTIF-----GLSSL 363
DB 310 PSDSILSINFGPTVDGDFLTDMPHTLLQLGKVKKAQLGVNKGDEGTAFLVYVAGPFSKD 369
QY 364 N--VTTNAQARAYFKQGFIIASDAEIDTLMAAYPDITQGSF-----FD--TGVLNALT 413
DB 370 NDSLITRKFQEGGLNMYFPFGVSRIGKRAVLFFVYVDWLGEQSPFVYRDALDDVDI 429
QY 414 P--QFKRISAVIGDLAFIHHARRYFLNHFGQGTYSFLSKOLSGLPIMGFTHANDIVWQDY 471
DB 430 PALEFTKKFALENNAPF---YFFEH-----RSSKLWPPEWGMVHGYEI---EF 473
QY 472 LLG---SGSVIYNNAFIATDL-----DPN-TAGLLVNNPKYTSSSQ 510
DB 474 VFGLPLRRVNYTRAEBIFRSIRINKMTWANFAKYGHGPNGTQGNSTMPVFTSTEQ 527


```

ranging from monomers and dimers to collagen-tailed and
hydrophobic-tailed forms, in which catalytic tetramers are
associated with anchoring proteins that attach them to the basal
lamina or to cell membranes. In the collagen-tailed forms, isoform
T subunits are associated with a specific collagen, COLQ, which
triggers the formation of isoform T tetramers, from monomers and
dimers.
```

-!- ALTERNATIVE PRODUCTS:

Event=Alternative splicing; Named isoforms=2;
Name=T;
IsoId=P23795-1; Sequence=Displayed;
Name=H;
IsoId=P23795-2; Sequence=VSP_001455;

-!- SIMILARITY: Belongs to the type-B carboxylesterase/lipase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>) or send an email to license@sib-sib.ch.

```

72 QY 72 SCMQ-----QNPEGTFEENLGTKALDVLVQSKVQFQAVLPQSEBCLTINVVRPPTKAGA 120
Db 99 -CYQYYVDTLVPFEGTEEMNPREL-----SEBCLYLN-VWTFYPRSS 140

126 QY 126 NLPVWLWIFGCGFEIGSPITFPFPAQVMTKSVLMGKHI-----LHVAVNYRVASWGFIAG 179
Db 141 PTPVLVWVYGGGYSGASSL-----DYYDGRFLTQABGTVLVSNMYRVGAFGLAL 191

180 QY 180 DDIAEBSGNAGLKQDLQMGOWADNIAGFGDPSKVTIIFGESAGSMVSLCHLIWNDGN 239
Db 192 PGSR-EAPGNVGLLDQRLALQWQENVAAFGGDPTSVTLFESAGAASVGMHLL-----244

240 QY 240 TYGKPLFRAGIMQSGAMVSDP---VDGTYGNEIYDLFVSSAGC-----GSASDKLAC 290
Db 245 SPFSRGLFRAVLQSGA--PNGPWATVGVGEARRRATLLARLVGCPGCGAGNDTEL VAC 302

291 QY 291 LRASSDTLDDATNTPGFLAYSRLSYLPRPDGKNIITDMYKLVDRGKYASVPVILGD 350
Db 303 LRARPAQDLVD--HEWRVLQESVFRFSFVPVVDGDFLSDTPEALINAGDFHGLQVLGV 360

351 QY 351 QNDEGTIF-----GLSSLN--VTTNQAQARYFKQSFIHASDAEIDTLMAY-----PQD 397
Db 361 VKDEGSYFLVYGAPGPKDNESLTSRAQFLAGVRVGPQASD LAEAAYVLHYTLWLHPED 420

398 QY 398 ---ITQSPDPTGVNLNALTPOFKRISAVLGDLAFIHARRYFLNHFQGGTKYSEL-----SK 450
Db 421 PARLREALSDVVGDNHVVCP-----VAQLAGSLA-----AQGARVYAITFEHRAS 465

451 QY 451 QLSGLPTWGTGFHANDIWDQYLLG---SGSVYNNAFIAFATDL-----DPNTA 496
Db 466 TLSWFLMWGVPHGYEI---EFIFGLPLEPSLNYITEERTFAQRLMYWANPARTGDPNDP 522

497 QY 497 --GLLVNMPKYTSSSQ 510
Db 523 RDPKAPQWPPYTAGAQ 538

Search completed: July 29, 2004, 10:49:33
Job time : 14 secs

```

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 10:47:09 ; Search time 21 Seconds
(without alignments)
2505.560 Million cell updates/sec

Title: US-09-943-857-4

Perfect score: 2864

Sequence: 1 SMNSRGPAGRLGVPKAKIA.....DNFRTAGYDALMTNPSPFFV 547

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

PIR 78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2779	97.0	549	JN0551	triacylglycerol li
2	2775	96.9	534	S41735	cholesterol estera
3	2513	87.7	549	S23448	triacylglycerol li
4	2431	84.9	549	JN0553	triacylglycerol li
5	2415	84.3	538	S05684	lipase I precursor
6	2366	82.6	549	JN0552	triacylglycerol li
7	2314	80.8	548	S32615	triacylglycerol li
8	1095	38.2	544	S41091	triacylglycerol li
9	1092	38.1	563	1 ACGUGC	triacylglycerol li
10	1089	38.0	544	S41093	triacylglycerol li
11	1086	37.9	544	S59958	triacylglycerol li
12	1084	37.8	544	S41092	triacylglycerol li
13	1081	37.7	544	S41096	triacylglycerol li
14	1081	37.7	544	S59957	triacylglycerol li
15	1078	37.6	563	1 PN0493	triacylglycerol li
16	1075	37.5	544	S41090	triacylglycerol li
17	1070	37.4	544	S41095	triacylglycerol li
18	1069	37.3	544	S41094	triacylglycerol li
19	491.5	17.2	596	1 ACRYE	acetylcholinestera
20	489	17.1	599	1 A38868	acetylcholinestera
21	481	16.8	614	2 JH0314	acetylcholinestera
22	476	16.6	603	2 S70849	cholinesterase (EC
23	469	16.4	614	2 JH0811	acetylcholinestera
24	461	16.1	583	2 S10712	acetylcholinestera
25	459	16.0	581	2 C39768	cholinesterase (EC
26	452	15.8	614	2 A39256	acetylcholinestera
27	451	15.7	584	2 T27009	hypothetical prote
28	451	15.7	602	1 ACHU	cholinesterase (EC
29	447.5	15.6	584	2 S48724	acetylcholinestera

ALIGNMENTS

RESULT 1

JN0551
triacylglycerol lipase (EC 3.1.1.3) 3 precursor - yeast (Candida rugosa)
N;Alternate names: lipase; triacylglycerol acylhydrolase
C;Species: Candida rugosa
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 07-May-1999
C;Accession: JN0551; S29633
R;Lotii, M.; Grandori, R.; Fusetti, F.; Longhi, S.; Brocca, S.; Tramontano, A.; Alberghi
Gene 124, 45-55, 1993
A;Title: Cloning and analysis of Candida cylindracea lipase sequences.
A;Reference number: JN0551; MUID:93178975; PMID:8440480
A;Accession: JN0551
A;Molecule type: DNA
A;Residues: 1-549 <LOT>
A;Cross-references: EMBL:X66008
A;Note: the source is designated as Candida cylindracea
R;Rua, M.L.; Diaz-Aurino, T.; Fernandez, V.M.; Otero, C.; Ballesteros, A.
Biochim. Biophys. Acta 1156, 181-189, 1993
A;Title: Purification and characterization of two distinct lipases from Candida cylindra
A;Reference number: S29633; MUID:93152596; PMID:8427877
A;Accession: S29633
A;Molecule type: protein
A;Residues: 16-25 <RUA>
A;Note: the source is designated as Candida cylindracea
C;Genetics:
A;Gene: LIP3
C;Function:
A;Description: hydrolyzes triacylglycerols into fatty acids and glycerol
C;Superfamily: triacylglycerol lipase
C;Keywords: carboxylic ester hydrolase; disulfide bond; glycoprotein
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-549/Product: triacylglycerol lipase 3 #status experimental <MAT>
F;222-226/Region: interfacial lipid recognition (GXSG) motif
F;75-112/Disulfide bonds: #status predicted
F;224/Active site: Ser #status predicted
F;329,366/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 97.0%; Score 2779; DB 2; Length 549;
Best Local Similarity 98.5%; Pred. No. 1.5e-192;
Matches 529; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY	11	LGSVPTAKLANGDTITGLNAINEAFGLIPFAEPVGNLRFKDPVPYSGSLNGKQKTSYG	70
DB	13	VAAAPTAKLANGDTITGLNAINEAFGLIPFAEPVGNLRFKDPVPYSGSLNGKQKTSYG	72
QY	71	PSCMQONPEGTTEENLGKLTALDLVMQSKVFQAVLPQSEDLITINVVRPPTKAGANLPVM	130
DB	73	PSCMQONPEGTTEENLGKLTALDLVMQSKVFQAVLPQSEDLITINVVRPPTKAGANLPVM	132
QY	131	LWIFGGFPIGSPITIPPAQMTKSVLMGKHHIVAVNRYVASWGLAGDDIKAECSGNA	190
DB	133	LWIFGGFPIGSPITIPPAQMTKSVLMGKHHIVAVNRYVASWGLAGDDIKAECSGNA	192

acetylcholinestera
para-nitrobenzyl e
acetylcholinestera
60K esterase (EC 3
carboxylesterase,
sterol esterase (E
carboxylesterase (E
carboxylesterase (E
acetylcholinestera
hypothetical prote
acetylcholinestera
acetylcholinestera
carboxylesterase (E
carboxylesterase (E
acetylcholinestera

30	442	15.4	607	2	T42399
31	430.5	15.0	489	2	B69680
32	427.5	14.9	620	2	A54413
33	423	14.8	532	2	A34329
34	418	14.6	540	2	A75250
35	411.5	14.4	597	2	A33668
36	410	14.3	565	2	S10367
37	409	14.3	549	2	JX0054
38	408.5	14.3	561	2	S47655
39	407	14.2	637	2	S66236
40	405.5	14.2	547	2	T29717
41	404.5	14.1	664	2	JC7990
42	404	14.1	691	2	JE0150
43	402	14.0	561	2	S62788
44	401	14.0	554	2	A39060
45	401	14.0	746	2	A25363

Qy	191	GLKDORLGMQWVADNIAGFGGDPSPKVTIIFGESAGSMVLCHLIWNWDGNTYTKGKELFRAG	250
Dd	193	GLKDORLGMQWVADNIAGFGGDPSPKVTIIFGESAGSMVLCHLIWNWDGNTYTKGKELFRAG	252
Qy	251	IMQSGAMVPSPDVGTYGNEYIDLFPVSSAGCGSASDKLACURSASDTLLDATNTPGFL	310
Dd	253	IMQSGAMVPSPDVGTYGNEYIDLFPVSSAGCGSASDKLACURSASDTLLDATNTPGFL	312
Qy	311	AYSLRLSYLPRPGKNITDDMYKLVRDGKYASVEVIIGDQNDEGTIFGLSLINVTTNAQ	370
Dd	313	AYSLRLSYLPRPGKNITDDMYKLVRDGKYASVEVIIGDQNDEGTIFGLSLINVTTNAQ	372
Qy	371	ARAYFKQSFIHASDAEDITLMAAPQDITQGSPEFTGVNLATPQFKRISAVLGDLATIH	430
Dd	373	ARAYFKQSFIHASDAEDITLMAAPQDITQGSPEFTGIFNAITPQFKRISAVLGDLATIH	432
Qy	431	ARRFYLNHFQGGTKYFSLSKQLSGLPIMGTFHANDIVMQDYLLGSGSVYNNAFATPATD	490
Dd	433	ARRFYLNHFQGGTKYFSLSKQLSGLPIMGTFHANDIVMQDYLLGSGSVYNNAFATPATD	492
Qy	491	LDPNTAGILLVNWPKYTTSSSQGNILMMINALGLYTGTKONFRTAGYDALMTNPSSFFV	547
Dd	493	LDPNTAGILLVNWPKYTTSSSQGNILMMINALGLYTGTKONFRTAGYDALMTNPSSFFV	549

```

RESULT 2
S41735
cholesterol esterase - yeast (Candida rugosa)
C:Species: Candida rugosa
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 17-Mar-1999
C:Accession: S41735
R:Kaiser, R.; Erman, M.; Duax, W.L.; Ghosh, D.; Joernvall, H.
FEBS Lett. 337, 123-127, 1994
A:Title: Monomeric and dimeric forms of cholesterol esterase from Candida cylindracea.
A:Reference number: S41735; MUID:94116663; PMID:8287964
A:Accession: S41735
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-534 <KAI>
A>Note: the source is designated as Candida cylindracea
C:Superfamily: triacylglycerol lipase

```

Query Match	96.9%;	Score 2775;	DB 2;	Length 534;
Best Local Similarity	99.1%;	Pred. No. 2.9e-192;		
Matches 528;	Conservative 3;	Mismatches 2;	Indels 0;	Gaps 0;
Qy	15	PTAKLANGDITITGLNAIINEAFILGIPAEAPPVGNLRKDPVPYSGSLNGOKFTSYGSPSCM	74	
Db	2	PTAKLANGDITITGLNAIINEAFILGIPAEAPPVGNLRKDPVPYSGSLNGOKFTSYGSPSCM	61	
Qy	75	QONPEGTFEENLKGKTALDLVQSKVFOAVLPQSEDCLTINVRPPGTKAGANLPMVLWIF	134	
Db	62	QONPEGTFEENLKGKTALDLVQSKVFOAVLPQSEDCLTINVRPPGTKAGANLPMVLWIF	121	
Qy	135	GGGFEIGSPITFPFAQMTKSVLMGKRIIHVAVNRYVASMGFLAGDDIKAEGSGNAGLKD	194	
Db	122	GGGFEIGSPITFPFAQMTKSVLMGKPIIHVAVNRYVASMGFLAGDDIKAEGSGNAGLKD	181	
Qy	195	QRLGMQWADNIAGFGGDPKVTITFGSAGSMSVLCHLIWNGDNTYKGKPLFRAGIMQS	254	
Db	182	QRLGMQWADNIAGFGGDPKVTITFGSAGSMSVLCHLIWNGDNTYKGKPLFRAGIMQS	241	
Qy	255	GAMVPSPDVGTYGNEIYDLFVSSAGCGSASDKLACLRASDSTLLLDATNNTPGFLAYSS	314	
Db	242	GAMVPSPDVGTYGNEIYDLFVSSAGCGSASDKLACLRASDSTLLLDATNNTPGFLAYSS	301	
Qy	315	LRLSYLPRPDGKNITDDMYKLVROGKYASVPVIIGQNDGTFELGSSLNVTWNAQARAY	374	
Db	302	LRLSYLPRPDGKNITDDMYKLVROGKYASVPVIIGQNDGTFELGSSLNVTWNAQARAY	361	
Qy	375	FKQSFIIHASDAEIDTLMAAPQDITQGSFPDGTGYNALITPQFKRISAVLGDFAFIHARRY	434	

362	FKQSGFIHSDAIEDITLMAYPQDITQGSFPDTGFNAITPQKRISAVLGDIAFHARY	421
Db		
435	FLNHFGQGTKYSPFSLKQSLGPIPMGTFTHANDIVWDYLLGSSSVTYNNAFIAFATDLDPN	494
Qy		
422	FLNHFGQGTKYSPFSLKQSLGPIPMGTFTHANDIVWDYLLGSSSVTYNNAFIAFATDLDPN	481
Db		
495	TAGLLVWNPKYTSSSQSGNNLMIMNALGLYTGKDNFTAGYDALMTNPSSFFV	547
Qy		
482	TAGLLVWNPKYTSSSQSGNNLMIMNALGLYTGKDNFTAGYDALMTNPSSFFV	534
Db		

RESULT 3
S23448

triacylglycerol lipase (EC 3.1.1.3) - yeast (*Candida rugosa*)
C/Species: *Candida rugosa*
C/Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 0
C/Accession: S23448
R/Longhi, S.; Fusetti, F.; Grandori, R.; Lotti, M.; Vanoni, R.
Biochim. Biophys. Acta 1131, 227-232, 1992
A/Title: Cloning and nucleotide sequences of two lipase genes
A/Reference number: S23448; PMID:1610906
A/Accession: S23448
A/Molecule type: DNA
A/Residues: 1-549 <LON>
A/Cross-references: EMBL:XG4703
A/Note: the source is designated as *Candida cylindracea*
C/Superfamily: triacylglycerol lipase
C/Keywords: carboxylic ester hydrolase

Query Match	87.7%	Score 2513;	DB 2;	Length 549;
Best Local Similarity	87.5%	Pred. No. 2.3e-173;		
Matches 470;	Conservative	32;	Mismatches 35;	Indels 0;
Gaps 0;				

Qy	11	LGSVPTAKLANGDITITGNAIINEAFLGIPRAEPVGNLRFKDPVPYSSGSLNGOKETS	70
Db	13	VAAPAPTAFANGDITITGNAIINEAFLGIPRAEPVGNLRFKDPVPYSSGSLGOKETS	72
Qy	71	PSCMQOONPEGTFFENLKGKTALDLYMQSKVQAVLQSEDCDCLTINVVRPGTTRAGANLP	130
Db	73	PSCMQOONPEGTFFENLPPAALDLYMQSKVFEAVSPSEDCDCLTINVVRPGTTRAGANLP	132
Qy	131	LWIEFGGQFEIGSPITFFPQAQVMVTKSVLKGKIIHVAIVNVRVASWGFLAGDDDIKAEGSGNA	190

[illegible]

RESULT 4
.TN0553

UN0553
triacylglycerol lipase (EC 3.1.1.3) 5 precursor - yeast (*Candida rugosa*)

N:Alternate names: lipase; triacylglycerol acylhydrolase
 C:Species: Candida rugosa
 C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 07-May-1999
 C:Accession: JN0553; S29634
 R:Lotti, M.; Grandori, R.; Fusetti, F.; Longhi, S.; Brocca, S.; Tramontano, A.; Alberghi
 Gene 124, 45-55, 1993
 A:Title: Cloning and analysis of Candida cylindracea lipase sequences.
 A:Reference number: JN0551; PMID:8440480
 A:Accession: JN0553
 A:Molecule type: DNA
 A:Residues: 1-549 <LOT>
 A:Cross-references: EMBL:X66008
 A:Note: the source is designated as Candida cylindracea
 R:Rua, M.L.; Diaz-Maurino, T.; Fernandez, V.M.; Otero, C.; Ballesteros, A.
 Biochim. Biophys. Acta 1156, 181-189, 1993
 A:Title: Purification and characterization of two distinct lipases from Candida cylindracea
 A:Reference number: S29633; PMID:8427877
 A:Accession: S29634
 A:Molecule type: protein
 A:Residues: 16-25 <RUA>
 A:Note: the source is designated as Candida cylindracea
 C:Genetics:
 A:Gene: LIP5
 C:Function:
 A:Description: hydrolyzes triacylglycerols into fatty acids and glycerol
 C:Superfamily: triacylglycerol lipase
 C:Keywords: carboxylic ester hydrolase; diacylhydride bond; glycoprotein
 F:1-15/Domain: signal sequence #status predicted <SIG>
 F:16-549/Product: triacylglycerol lipase 5 #status experimental <MAT>
 F:222-226/Region: interfacial lipid recognition (GXSGX) motif
 F:75-112/Diulfide bond: #status predicted
 F:224/Active site: Ser #status predicted
 F:329,366/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 84.9%; Score 2431; DB 2; Length 549;
 Best Local Similarity 85.5%; Pred. No. 1.9e-167;
 Matches 459; Conservative 32; Mismatches 46; Indels 0; Gaps 0;

QY	11	LGSVPTAKLANGDTITGLNAINEAFGLIPFAEPVGNLRFKDPVYSGSLNGKQTSYG	70
DB	13	VAAAPTATLANGDTITGLNAINEAFGLIPFAEPVGNLRFKDPVYSGSLNGKQTSYAG	72
QY	71	PSCMQONPEGTPEENLGKTDALDVMSQKQVAVLPQSECLTINVRPPTKAGANLPVM	130
DB	73	PSCMQONPEGTPEENLPKVALDLVMSQKQVAVLPQSECLTINVRPPTKAGANLPVM	132
QY	131	LWIFGGFEIGSPITFPPAQMTKSVLMGKHIIHVAVNRYVSWGFLAGDDIIKAEGSGNA	190
DB	133	LWIFGGFEIGSPITFPPAQMTKSVLMGKHIIHVAVNRYVSWGFLAGDDIIKAEGSSNA	192
QY	191	GLKQDRGLQMWADNITAGFGDPSKVTIFGESAGSMVLCILWNGDNTYKGPFLFRAG	250
DB	193	GLKQDRGLQMWADNITAGFGDPSKVTIFGESAGSMVLCILWNGDNTYKGPFLFRAG	252
QY	251	IMQSGAMVPSPDVGTYGNEIYDLFVSSAGCGSASDKLACLRASSDITLLDATNTPGFL	310
DB	253	IMQSGAMVPSPDVGTYGTYQIYDLFVASTCGSSASDKLACLRASSDITLLDATNTPGFL	312
QY	311	AYSSLRLSYLPRDGNITDDMYKLVDRDGKYASVPVIGQNDGEGTIFGLSSLNVTNNAQ	370
DB	313	SYTSLRLSYLPRDGNITDDMYKLVDRDGKYASVPVIGQNDGEGTIFGLSSLNVTTEAD	372
QY	371	ARAYFKQSFTHASDAEIDTLMAAYPDITQSGPDTGVNALTPOPKRISAVLGDLAFIH	430
DB	373	AEAYLRKSFTHATDADITALKAAAYPSDVTQSGPDTGILNALTPOPKRINAVLGDLFTTL	432
QY	431	ARRYFLNHFGGTYKYSFLSKQLSGLPIMGTFFHNDIVWQDYLLGSGSVIYNNAFIAFATD	490
DB	433	SRRFLNHYTGTPKYSFLSKQLSGLPILGTFFHNDIVWQDYLLGSGSVIYNNAFIAFATD	492
QY	491	LDPNTAGLLVNPKYTSQSGNNLMINAGLYTGKDNFTAGYDALMTNPSPFFV	547
DB	493	LDPNTAGLLVNPKYTSQSGAGNNLMQISALGLYTGKDNFTAGYDALMTNPSPFFV	549

RESULT 5

S05684
 lipase I precursor - yeast (Candida rugosa) (fragment)
 C:Species: Candida rugosa
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 13-Sep-1998
 C:Accession: S05684
 R:Kawaguchi, Y.; Honda, H.; Taniguchi-Morimura, J.; Iwasaki, S.
 Nature 341, 164-166, 1989
 A:Title: The codon CUG is read as serine in an asporogenic yeast Candida cylindracea.
 A:Reference number: S05684; PMID:89384874; PMID:2506450
 A:Accession: S05684
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-538 <KAW>
 A:Note: part of this sequence, including the amino end of the mature protein, was confir
 A:Note: the source is designated as Candida cylindracea
 C:Superfamily: triacylglycerol lipase
 F:1-4/Domain: signal sequence (fragment) #status predicted <SIG>
 F:5-538/Product: lipase I #status experimental <MAT>

Query Match 84.3%; Score 2415; DB 2; Length 538;
 Best Local Similarity 84.5%; Pred. No. 2.7e-166;
 Matches 454; Conservative 32; Mismatches 51; Indels 0; Gaps 0;

QY	11	LGSVPTAKLANGDTITGLNAINEAFGLIPFAEPVGNLRFKDPVYSGSLNGKQTSYG	70
DB	2	VAAAPTATLANGDTITGLNAINEAFGLIPFAEPVGNLRFKDPVYSGSLNGKQTSYG	61
QY	71	PSCMQONPEGTPEENLGKTDALDVMSQKQVAVLPQSECLTINVRPPTKAGANLPVM	130
DB	62	PLCMQONPEGTPEENLPKVALDLVMSQKQVAVLPQSECLTINVRPPTKAGANLPVM	121
QY	131	LWIFGGFEIGSPITFPPAQMTKSVLMGKHIIHVAVNRYVSWGFLAGDDIIKAEGSGNA	190
DB	122	LWIFGGFEIGSPITFPPAQMTKSVLMGKHIIHVAVNRYVSWGFLAGDDIIKAEGSANA	181
QY	191	GLKQDRGLQMWADNITAGFGDPSKVTIFGESAGSMVLCILWNGDNTYKGPFLFRAG	250
DB	182	GLKQDRGLQMWADNITAGFGDPSKVTIFGESAGSMVLCILWNGDNTYKGPFLFRAG	241
QY	251	IMQSGAMVPSPDVGTYGNEIYDLFVSSAGCGSASDKLACLRASSDITLLDATNTPGFL	310
DB	242	IMQSGAMVPSPDVGTYGNEIYDLFVSSAGCGSASDKLACLRASSDITLLDATNTPGFL	301
QY	311	AYSSLRLSYLPRDGNITDDMYKLVDRDGKYASVPVIGQNDGEGTIFGLSSLNVTNNAQ	370
DB	302	AYSSLRLSYLPRDGNITDDMYKLVDRDGKYASVPVIGQNDGEGTIFGLSSLNVTTDAQ	361
QY	371	ARAYFKQSFTHASDAEIDTLMAAYPDITQSGPDTGVNALTPOPKRISAVLGDLAFIH	430
DB	362	ARAYFKQSFTHASDAEIDTLMAAYPDITQSGPDTGVNALTPOPKRISAVLGDLGFTL	421
QY	431	ARRYFLNHFGGTYKYSFLSKQLSGLPIMGTFFHNDIVWQDYLLGSGSVIYNNAFIAFATD	490
DB	422	ARRYFLNHYTGTPKYSFLSKQLSGLPVLTFFHNDIVWQDYLLGSGSVIYNNAFIAFATD	481
QY	491	LDPNTAGLLVNPKYTSQSGNNLMINAGLYTGKDNFTAGYDALMTNPSPFFV	547
DB	482	LDPNTAGLLVNPKYTSQSGNNLMINAGLYTGKDNFTAGYDALMTNPSPFFV	538

RESULT 6

JN0552
 triacylglycerol lipase (EC 3.1.1.3) 4 precursor - yeast (Candida rugosa)
 N:Alternate names: lipase; triacylglycerol acylhydrolase
 C:Species: Candida rugosa
 C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 07-May-1999
 C:Accession: JN0552
 R:Lotti, M.; Grandori, R.; Fusetti, F.; Longhi, S.; Brocca, S.; Tramontano, A.; Alberghi
 Gene 124, 45-55, 1993
 A:Title: Cloning and analysis of Candida cylindracea lipase sequences.


```

Best Local Similarity 43.5%; Pred. No. 4.1e-71;
Matches 238; Conservative 72; Mismatches 193; Indels 44; Gaps 9;

QY 15 PTAKLANGTITGLNAINAEFLGIPPAEPVGNLRFKDPVPYSGSLNGQKTSYCPSCM 74
Db 3 PTAVLNGNEVISGLVGKVDTFKGIFFADPPVGDLPFKHPQPTGSGYQGLKANDFSSACM 62
QY 75 QNPECTFE-----ENLCKTALDLVMQSKVFQAVLPQSEDCLTINVRPPGT 121
Db 63 QLDPGNAISLLDKVGLGKILPDNLGRPLYDMA-----QGSVMNEDCLYLVFPAGT 116
QY 122 KGANLPVMLWIFGGGFEIGSPITFPPAQMVTKSVLMGKHIIHVAVNRYVASVLAGDD 181
Db 117 KPDALPVMVWIYGGAFVFGSSASYPGNGVYKESVEMGQPVVVFVSINRYTGPYGLCGDA 176
QY 182 IKAEGSGNAGLKQRLGMOWADNIAGFGDPSKVTIFGESAGSMVYLCHLIWNGDNTY 241
Db 177 ITAEGSTNAGLHQRLGLEWVSDNIANFGGDPDKVMIFGESAGMSVAHQLVAYGSDNTY 236
QY 242 KGKPLFRAGIMQSGAMVPSDPVDTGYNEIYDLFVSSAGC---GSASDDKLACLRSSSDT 298
Db 237 NGQLFHSAIQSGGFLPFDSTSVGPESAYSRAQYAGCDTSASDNDTLACLRSSSDV 296
QY 299 LLDATNN-----TPGFLAYSSRLSYLPRPDGKNITDDMYKLVRDGKYASVPIIG 349
Db 297 LHSQNSYDLKDLFGLLPQFLGFG-----PRPDGNIIPDAAYELVRSGRYAKVPIIG 349
QY 350 QNDDEGTIFGLSLNVTNNAQARAYFKQSFHASDAEIDTLMAAYPQDITQGSFPFTGVL 409
Db 350 NOEDEGITLAPVAINATTTPHVKKWLKICQSQASDASLDRVLSFGSSEGSFPFTGIL 409
QY 410 NALTPQFKRISAVGLDLPATFHARRYFLNHFQGGTKYFSLSKQLSGL-PIMGTFHNDIVM 468
Db 410 NALTPQFKRIIAFTLLFQSPRVMLNATKDVNRVTLATQLNLVLPFTGTHGSDLLF 469
QY 469 QDYL-LGSGSVIYNNAFIATPDLDPNTAGLLVNWPKYTSSSQSGNNLMINALGTYGK 527
Db 470 QYVVDLGPSA-YRRYFISFANHHDPNVTNLQQWDMYT---DAGKEMLIQHMGNSMRT 525
QY 528 DNFRTAG 534
Db 526 DDFRIEG 532

RESULT 9
ACGUCG
triacylglycerol lipase (EC 3.1.1.3) I precursor - yeast (Geotrichum candidum)
N;Alternate names: lipase
C;Species: Geotrichum candidum
C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 31-Mar-2000
C;Accession: P0492; JQ0022
R;Nagao, T.; Shimada, Y.; Sugihara, A.; Tominaga, Y.
J. Biochem. 113, 776-780, 1993
A;Title: Cloning and sequencing of two chromosomal lipase genes from Geotrichum candidum
A;Reference number: P0492; MUID:93380907; PMID:8370674
A;Accession: P0492
A;Molecule type: DNA
A;Residues: 1-563 <NAG>
A;Note: The translation of residues 31-550 and the corresponding nucleotide sequence are
R;Shimada, Y.; Sugihara, A.; Tominaga, Y.; Iizumi, T.; Tsunawawa, S.
J. Biochem. 106, 383-388, 1989
A;Title: cDNA molecular cloning of Geotrichum candidum lipase.
A;Reference number: JQ0022; MUID:90110016; PMID:2481674
A;Accession: JQ0022
A;Molecule type: mRNA
A;Residues: 1-563 <SHI>
A;Experimental source: strain ATCC 34614
A;Note: sequences of several small peptides were also determined
C;Comment: The extracellular lipase produced by Geotrichum candidum hydrolyzes all ester
C;Genetics:
A;Gene: lipi
C;Superfamily: cholinesterase; cholinesterase homology
C;Keywords: carboxylic ester hydrolase; glycoprotein; pyroglutamic acid

```

```

F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-563/Product: triacylglycerol lipase #status experimental <MAT>
F:43-560/Domain: cholinesterase homology <CHE>
F:234-238/Region: interfacial lipid recognition (GXSG) motif
F:20/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimen
F:80-124,295-307/Disulfide bonds: #status predicted
F:236/Active site: Ser #status predicted
F:302,383/Binding site: carbohydtrate (Asn) (covalent) #status predicted

Query Match 38.1%; Score 1092; DB 1; Length 563;
Best Local Similarity 42.9%; Pred. No. 7.1e-71;
Matches 237; Conservative 72; Mismatches 201; Indels 42; Gaps 8;

QY 9 GRIGSVPTAKLANGDITITGLNAINAEFLGIPPAEPVGNLRFKDPVPYSGSLNGQKTS 68
Db 16 GTLAQAPTAVLNGNEVISGLVGKVDTFKGIFFADPPVGDLPFKHPQPTGSGYQGLKAND 75
QY 69 YGSPCMQONPEGTFE-----ENLCKTALDLVMQSKVFQAVLPQSEDCLTINV 115
Db 76 FSSACMLQDFGNAISLLDKVGLGKILPDNLGRPLYDMA-----QGSVMNEDCLYLV 129
QY 116 VPPGPTKAGANLPVMLWIFGGGFEIGSPITFPPAQMVTKSVLMGKHIIHVAVNRYVASWG 175
Db 130 FRAGTKPDAKLPVMVWIYGGAFVFGSSASYPGNGVYKESLEMGQPVVVFVSINRYTGPY 189
QY 176 FLAGDDIKAEGSGNAGLKQRLGMOWADNIAGFGDPSKVTIFGESAGSMVYLCHLIWN 235
Db 190 FLGGDAITAEQNTNAGLHQRLGLEWVSDNIANFGGDPDKVMIFGESAGMSVAHQLVAY 249
QY 236 DGNTYKGPFLFRAGIMQSGAMVPSDPVDTGYNEIYDLFVSSAGC-GSASDK--LACL 292
Db 250 GGDNTYNGQLFHSAIQSGGFLPFDSTSVGPESAYSRAQYAGCDASAGDNETLACL 309
QY 293 SASDITLLDATNN-----TPGFLAYSSRLSYLPRPDGKNITDDMYKLVRDGKYAS 343
Db 310 SKSDVLHSAQNSYDLKDLFGLLPQFLGFG-----PRPDGNIIPDAAYELVRSGEYAK 362
QY 344 VPVIIGQNDDEGTIFGLSLNVTNNAQARAYFKQSFHASDAEIDTLMAAYPQDITQGS 403
Db 363 VPVIIGQNDDEGTIFGLSLNVTNNAQARAYFKQSFHASDAEIDTLMAAYPQDITQGS 422
QY 404 FDTGVNALTTPQFKRISAVGLDLPATFHARRYFLNHFQGGTKYFSLSKQLSGL-PIMGTF 462
Db 423 FRTGILNALTTPQFKRIIAFTLLFQSPRVMLNATKDVNRVTLATQLNLVLPFTGTH 482
QY 463 ANDIVMQDYLGGSGSVIYNNAFIATPDLDPNTAGLLVNWPKYTSSSQSGNNLMINALG 522
Db 483 GSDLLFYVYAGPWSSSAYRRYFISFANHHDPNVTNLQQWDMYT---DSGKEMLIQH 539
QY 523 LYTGDKNFRTAG 534
Db 540 NSMRTDDFRIEG 551

RESULT 10
S41093
triacylglycerol lipase (EC 3.1.1.3) I precursor - yeast (Geotrichum candidum) (strain NF
C;Species: Geotrichum candidum
A;Variety: NRRL Y-553
C;Date: 19-Mar-1997 #sequence_revision 05-Feb-1999 #text_change 18-Jun-1999
C;Accession: S41093
R;Bertolini, M.C.; Laramee, L.; Thomas, D.Y.; Cygler, M.; Schrag, J.D.; Vernet, T.
Eur. J. Biochem. 219, 119-125, 1994
A;Title: Polymorphism in the lipase genes of Geotrichum candidum strains.
A;Reference number: S41090; MUID:94139683; PMID:8306978
A;Accession: S41093
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-544 <BER>
A;Cross-references: GB:U02387; NID:9406507; PIDN:AAA03425.1; PID:9406508
A;Experimental source: NRRL Y-553
A;Note: only the translation of the mature protein is shown
C;Genetics:

```

A:Gene: lipi

C:Function:

A:Description: hydrolyzes triacylglycerols into fatty acids and glycerol

C:Superfamily: cholinesterase; cholinesterase homology

C:Keywords: carboxylic ester hydrolase; glycoprotein; lipid hydrolysis; pyroglutamic acid

F:215-541/Domain: cholinesterase homology <CHE>

F:1/Modified site: interfacial lipid recognition (GXSGX) motif

F:61-105,276-288/disulfide bonds: #status predicted

F:217/Active site: Ser #status predicted

F:283,364/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 38.0%; Score 1089; DB 2; Length 544;
Best Local Similarity 43.5%; Pred. No. 1.1e-70;
Matches 238; Conservative 73; Mismatches 192; Indels 44; Gaps 10;

```
QY 15 PTAKLANGDTITGLNAIINEAFGLIPFAEPVGNLRFKDPVPYSGSLNGQKFTSYGSPSCM 74
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 3 PTAVLNGNEVISGLVGKVDTFKGIIPADPPVGLDRFKHPQFTGSYQGLKANDFFSSACM 62
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 75 QONPEGTPE-----ENLGKLTALDLVMQSKVFAVLPOSEDCLTINNVRRPPT 121
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 63 QLDPFNAISLWDKVVGLGKILPDNLRGPLYDMA-----QGSVSNEDCLINVRFPAGT 116
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 122 KAGANLPVWLMTFGGFGFPGISPTIPPPAQMTKSVLMGKHIHVAVNYRVASWGLAGDD 181
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 117 KPDAKLPVWMIYGGAFVFGSSASYPGNGYVKESVEMGQPVVVFVSINRTGPGYFLGGDA 176
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 182 IKAEGSGNAGLKQDLGCMQWADNIAGFGDPSPKVTIFGESAGSMVSLCHLIWNGDNTY 241
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 177 ITAEGNTNAGLHDQKGLGWSDNIANFGGDPDKVMIFGESAGAMSAVHQLVAYGDDNTY 236
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 242 KGKPLFRAGIMQSGAMVPSDPVDGTGYGNEIYDLFVSSAGC--GSASDK--LACLRASSDT 298
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 237 NGKKLFHSAIILQSGGLPYFDSTSVGPESASRFAQYAGCDASDNETLACLRKSSDV 296
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 299 LLDATNN-----TPGFAYSSLRLSYLPRPGKNITDDMYKLVRDGVKVASVPVITG 349
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 297 LHSAQNSYDLKDLGLLPOQLGFG-----PRPDGNIIPDAAYELRSGRYAKVPYITG 349
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 350 DONDEGTIFGLSSLVNTTNAQARAYFKQSFHASDAEITDLMAAYPQDITOGSPDPTGVL 409
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 350 NODEGTILAPVAINATTPHVKWLVKICSEASASLDRVLSLYPGSWSEGAPRTGIL 409
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 410 NALTPOPKRISAVLGDALAFIHARRYFLNHFGGKTKYSPLSKQLSGL--PIMGTFHNDIVW 468
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 410 NALTPOPKRIAAITDILLFQSPRVRVNLNATKDVNRTYLATQLHNLVFLGTFFHGSDDLIF 469
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 469 QDYL-LGSGSVIYNNAFIATATDLPNTAGLLVNWPKYTSQSGNNLMINALGLYTGK 527
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 470 QYYVDLGPSSA--YRRYFISFANHHDPNVGTLNQWDMYT---DAGKEMLIQIHVMVNSMRT 525
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 528 DNERTAG 534
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 526 DDFRIEG 532
```

RESULT 11

S59958

triacylglycerol lipase (EC 3.1.1.3) I - yeast (Geotrichum candidum) (strain CBS 178.71)

C:Species: Geotrichum candidum

A:Variety: strain CBS 178.71

C:Date: 09-Mar-1996 #sequence_revision 19-Apr-1996 #text_change 21-Jan-2000

C:Accession: S59958; S49236

R:Phillips, A.; Pretorius, G.H.J.; van Rensburg, H.G.

Biochim. Biophys. Acta 1252, 305-311, 1995

A:Title: Molecular characterization of a Galactomyces geotrichum lipase, another member

A:Reference number: S59957; MUID:96049515; PMID:7578238

A:Accession: S59958

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-544 <PH>

A:Cross-references: EMBL:X81656

A:Experimental source: strain CBS 178.71

R:Phillips, A.; Pretorius, G.H.J.; van Rensburg, H.G.

submitted to the EMBL Data Library, September 1994

A:Reference number: S49236

A:Accession: S49236

A:Molecule type: DNA

A:Residues: 1-82, 'L', '84-239, 'Q', '241-383, 'Y', '385-544 <PHW>

A:Cross-references: EMBL:X81656; NID:g547503; PIDN:CAA57316.1; PID:g547504

C:Genetics:

C:Gene: lipi

C:Superfamily: cholinesterase; cholinesterase homology

C:Keywords: carboxylic ester hydrolase

F:24-541/Domain: cholinesterase homology <CHE>

Query Match 37.9%; Score 1086; DB 2; Length 544;

Best Local Similarity 43.1%; Pred. No. 1.8e-70;

Matches 236; Conservative 73; Mismatches 194; Indels 44; Gaps 9;

```
QY 15 PTAKLANGDTITGLNAIINEAFGLIPFAEPVGNLRFKDPVPYSGSLNGQKFTSYGSPSCM 74
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 3 PTAVLNGNEVISGLVGKVDTFKGIIPADPPVGLDRFKHPQFTGSYQGLKANDFFSSACM 62
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 75 QONPEGTPE-----ENLGKLTALDLVMQSKVFAVLPOSEDCLTINNVRRPPT 121
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 63 QLDPFNAISLWDKVVGLGKILPDNLRGPLYDMA-----QGSVSNEDCLINVRFPAGT 116
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 122 KAGANLPVWLMTFGGFGFPGISPTIPPPAQMTKSVLMGKHIHVAVNYRVASWGLAGDD 181
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 117 KPDAKLPVWMIYGGAFVFGSSASYPGNGYVKESVEMGQPVVVFVSINRTGPGYFLGGDA 176
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 182 IKAEGSGNAGLKQDLGCMQWADNIAGFGDPSPKVTIFGESAGSMVSLCHLIWNGDNTY 241
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 177 ITAEGNTNAGLHDQKGLGWSDNIANFGGDPDKVMIFGESAGAMSAVHQLVAYGDDNTY 236
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 242 KGKPLFRAGIMQSGAMVPSDPVDGTGYGNEIYDLFVSSAGC---SASKLACLRASSDT 298
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 237 NGKKLFHSAIILQSGGLPYFDSTSVGPESASRFAQYAGCDTSDNDTLACLRKSSDV 296
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 299 LLDATNN-----TPGFAYSSLRLSYLPRPGKNITDDMYKLVRDGVKVASVPVITG 349
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 297 LHSAQNSYDLKDLGLLPOQLGFG-----PRPDGNIIPDAAYELRSGRYAKVPYITG 349
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 350 DONDEGTIFGLSSLVNTTNAQARAYFKQSFHASDAEITDLMAAYPQDITOGSPDPTGVL 409
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 350 NODEGTILAPVAINATTPHVKWLVKICSEASASLDRVLSLYPGSWSEGAPRTGIL 409
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 410 NALTPOPKRISAVLGDALAFIHARRYFLNHFGGKTKYSPLSKQLSGL--PIMGTFHNDIVW 468
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 410 NALTPOPKRIAAITDILLFQSPRVRVNLNATKDVNRTYLATQLHNLVFLGTFFHGSDDLIF 469
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 469 QDYL-LGSGSVIYNNAFIATATDLPNTAGLLVNWPKYTSQSGNNLMINALGLYTGK 527
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 470 QYYVDLGPSSA--YRRYFISFANHHDPNVGTLNQWDMYT---DAGKEMLIQIHVMVNSMRT 525
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 528 DNERTAG 534
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 526 DDFRIEG 532
```

RESULT 12

S41092

triacylglycerol lipase (EC 3.1.1.3) I precursor - yeast (Geotrichum candidum) (strain

C:Species: Geotrichum candidum

A:Variety: NRRL Y-552

C:Date: 19-Mar-1997 #sequence_revision 05-Feb-1999 #text_change 18-Jun-1999

C:Accession: S41092

R:Bertolini, M.C.; Laramee, L.; Thomas, D.Y.; Cygler, M.; Schrag, J.D.; Vernet, T.

Eur. J. Biochem. 219, 119-125, 1994

A:Title: Polymorphism in the lipase genes of Geotrichum candidum strains.

A:Reference number: S41090; MUID:94139683; PMID:8306978

A:Accession: S41092

A:Status: nucleic acid sequence not shown; not compared with conceptual translation

A:Molecule type: DNA

A;Title: Molecular characterization of a Galactomyces geotrichum lipase, another member
A;Reference number: S59957; MUID:96049515; PMID:7578238
A;Accession: S59957
A;Molecule type: nucleic acid
A;Residues: 1-563 <PH>
A;Cross-references: EMBL:X78032; NID:9459249; PIDN:CAA54963.1; PID:9459250
C;Superfamily: cholinesterase; cholinesterase homology
C;Keywords: carboxylic ester hydrolase
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-563/Product: triacylglycerol lipase #status predicted <MAT>
F;43-560/Domain: cholinesterase homology <CH>
Query Match 37.7%; Score 1081; DB 2; Length 563;
Best Local Similarity 43.7%; Pred. No. 4.4e-70;
Matches 240; Conservative 68; Mismatches 199; Indels 42; Gaps 8;
Qy 9 GRLGSVPTAKLANGDTITGLNAINAEPLGTPFAEPVGNLRFKDPVPYSGSLNGQKRTS 68
Db 16 GALAQTPFAVLNVNNOVIGVLEKGVDTFKGIPFGAPPVGDRLRFKHPQPFQFTGYQGLKAND 75
Qy 69 YGPSQMOQNPGETF-----EENLGKTLALDVMQSKVFOAVLPOSEDCLTINV 115
Db 76 FSSACMQLNPNALNTILDNALSLASISPENIRGLYDMAKGS-----VMSSEDCLYLNV 129
Qy 116 VRPGTKAGANLPMVLMTFGGFGIGSTPIPPAQMTKSVLMGKHIIHVAVNRYVASWG 175
Db 130 CRPAGTKPGDKLPVMWMIYGAFTFGSSRYPGNGYVDESVKMGQPVVVFVSINVRSGBYG 189
Qy 176 FLAGDDIKAEKSGNAGLKDOPLGMQWADNITAGGDPSPKVTIFGESAGSMVLCILWN 235
Db 190 FLGGDGHITAEAGNTNAGLDQRLGLEWSDNIANFGGDPCKVMIIFGESAGSMVGHLLTAY 249
Qy 236 DGDNTYKGPFLFRAGIMQSGAMVPSDPVDGTGNYGNEIYDLFVSSAGC--SASDK--LACLR 292
Db 250 GGDNTYNGKALFHSAILOQSGPLPYNSGWLGPDSAYNRFAYAGCDTSASDVILQCLR 309
Qy 293 SASSDTLIDATNN-----TPGFLAYSSRLSYLPRPGKMTIDDMYKLVLDGKYAS 343
Db 310 SKPSTSLSDAQSNDLKDGLFGLLPQFLGFG-----PRPDGDIIPDSAYELVRSGRYAK 362
Qy 344 VPVIIGQNDGFTIFGLSSLVNVTNAQARAYFKQSFHASDAEIDTLMAAYPQDITQSP 403
Db 363 VPYITGQDEGTILAPVAINATTTPHVKWKLKIFNEATDSLDRVLKLPETLSEGS 422
Qy 404 FDTGLVNALTPOFKRISAVLGDALFIHARRYFLNHFQGGTKYFSLSKQLSGL--PIMGTFH 462
Db 423 FRTGLVNALTPOFKRVAATFDLLFQSPRDVWLNATKDQVNEWTFLATQLHNLVFFLGTFF 482
Qy 463 ANDIVQDYLGGSGSVIYNNAFATDLDNTAGLLVNNPKYTSQSSQSGNNLMMLNALG 522
Db 483 ASDVLFQYLYNIGPSDSYLYFISFRNHDPNVGTGLQWAKYT---DGGKEMLEIKMLG 539
Qy 523 LYTCKDNFR 531
Db 540 NSMRTDDFR 548
RESULT 15
PN0493
C;Species: Geotrichum candidum
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 16-Jun-2000
C;Accession: PN0493; A46760
R;Nagao, T.; Shimada, Y.; Sugihara, A.; Tominaga, Y.
J. Biochem. 113, 776-780, 1993
A;Title: Cloning and sequencing of two chromosomal lipase genes from Geotrichum candidum
A;Reference number: PN0492; MUID:93380907; PMID:8370674
A;Accession: PN0493
A;Molecule type: DNA
A;Residues: 1-563 <NAG>
A;Note: the translation of residues 31-550 and the corresponding nucleotide sequence are
R;Shimada, Y.; Sugihara, A.; Iizumi, T.; Tominaga, Y.
J. Biochem. 107, 703-707, 1990

A;Title: cDNA cloning and characterization of Geotrichum candidum lipase II.
A;Reference number: A46760; MUID:90375435; PMID:2398037
A;Accession: A46760
A;Molecule type: mRNA
A;Residues: 7-563 <SHI>
A;Cross-references: GB:D00697; NID:g217926; PIDN:BAA00603.1; PID:g217927
C;Comment: The extracellular lipase produced by Geotrichum candidum hydrolyzes all este
C;Genetics:
A;Gene: lipII
C;Superfamily: cholinesterase; cholinesterase homology
C;Keywords: carboxylic ester hydrolase; glycoprotein; pyroglyutamic acid
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-563/Product: triacylglycerol lipase II #status predicted <MAT>
F;43-560/Domain: cholinesterase homology <CH>
F;234-238/Region: interfacial lipid recognition (GXSGXG) motif
F;20/Modified site: pyroglutamic acid (Gln) (in mature form) #status predicted
F;80-124,235-307/Disulfide bonds: #status predicted
F;236/Active site: Ser #status predicted
F;302,383/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 37.6%; Score 1078; DB 1; Length 563;
Best Local Similarity 41.6%; Pred. No. 7.2e-70;
Matches 233; Conservative 83; Mismatches 202; Indels 42; Gaps 8;
Qy 8 AGRILGSVPTAKLANGDTITGLNAINAEPLGIPFAEPVGNLRFKDPVPYSGSLNGQKFT 67
Db 15 AGVLAQAPRPSPSLNGNEVISGLVGKVDTFKGIIPADPLNDRFKHPQPFQFTGSGYQGLKAN 74
Qy 68 SYGFSQMOQNPGETFEEENLGKTLALDLM--QSKVF-----QAVLPOSEDCLTIN 114
Db 75 DFSACMQLDPGNSL-----TLIDKALGLAKVIFEEFRGLPYDMAKGTVMSEDCLYIN 128
Qy 115 VVRPGTKAGANLPMVLMTFGGFGIGSTPIPPAQMTKSVLMGKHIIHVAVNRYVASW 174
Db 129 VFRPAGTKPAKLPMWMIYGAFTFGSSRAYPGNSYKVSINMGQPVFVSINRTGPF 188
Qy 175 GFLAGDDIKAEKSGNAGLKDOPLGMQWADNITAGGDPSPKVTIFGESAGSMVLCILW 234
Db 189 GFLGGDAITAEAGNTNAGLDQRLGLEWSDNIANFGGDPDKVMIIFGESAGSMVAHQLIA 248
Qy 235 NDGNTYKGPFLFRAGIMQSGAMVPSDPVDGTGNYGNEIYDLFVSSAGC---GSASDKLACL 291
Db 249 YGGDNTYNGKKLFHSAILOQSGPLPYHDSVSGPDISYNRFAQYAGCDTSASANDTLECL 308
Qy 292 RSASDSTLLIDATNN-----TPGFLAYSSRLSYLPRPGKMTIDDMYKLVLDGKYA 342
Db 309 RSKSSSVLHDAQNSYDLKDLFGLLPQFLGFG-----PRPDGNIIPDAAYELFRSGRYA 361
Qy 343 SVPVIIGQNDGFTIFGLSSLVNVTNAQARAYFKQSFHASDAEIDTLMAAYPQDITQGS 402
Db 362 KVPYISGNQDEGTAFAPVALNATTPHVKWKLQYIFVDASEASIDRVLSLYPQTLSVGS 421
Qy 403 PFDGLVNALTPOFKRISAVLGDALFIHARRYFLNHFQGGTKYFSLSKQLSGL--PIMGTF 461
Db 422 PFTGILVNALTPOFKRVAAILSDMLFQSPRRVMLSATKDQVNRVYLSLTHLNLVFFLGTFF 481
Qy 462 HANDIVQDYLGGSGSVIYNNAFATDLDNTAGLLVNNPKYTSQSSQSGNNLMMLNAL 521
Db 482 HGNELIFQFNVTGPNPANSYLYRFLISFANHHDPNVGTNLLQWDQYT---DEGKEMLEIHTM 538
Qy 522 GLYTGKDNFRTAGYDALMTN 541
Db 539 DNVMTDDYRIEGISNPFETD 558
Search completed: July 29, 2004, 10:51:00
Job time : 23 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 10:50:34 ; Search time 50 Seconds
(without alignments)
3431.691 Million cell updates/sec

Title: US-09-943-857-4
Perfect score: 2864
Sequence: 1 SMNSRGPAGRLGVSPTAKLA.....DNFRTAGYDALMTNPSFFV 547

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2864	100.0	547	10	US-09-943-857-4
2	2588	90.4	547	10	US-09-943-857-10
3	2495	87.1	547	10	US-09-943-857-8
4	2434	85.0	547	10	US-09-943-857-6
5	2396	83.7	547	10	US-09-943-857-2
6	1170.5	40.9	561	15	US-10-369-493-4236
7	674.5	23.6	612	9	US-09-875-353-4
8	672.5	23.5	574	13	US-10-023-515-4
9	672.5	23.5	574	16	US-10-674-636-4
10	664	23.2	585	9	US-09-934-323-4
11	482	16.8	574	9	US-09-748-739A-21
12	482	16.8	574	14	US-10-032-233-48
13	482	16.8	574	16	US-10-324-466-48
14	482	16.8	574	16	US-10-413-432-48
15	477	16.7	574	9	US-09-748-739A-23

16	477	16.7	574	14	US-10-032-233-50	Sequence 50, Appl
17	477	16.7	574	16	US-10-324-466-50	Sequence 50, Appl
18	477	16.7	574	16	US-10-413-432-50	Sequence 50, Appl
19	459	16.0	574	9	US-09-748-739A-22	Sequence 22, Appl
20	459	16.0	574	14	US-10-032-233-49	Sequence 49, Appl
21	459	16.0	574	16	US-10-324-466-49	Sequence 49, Appl
22	459	16.0	574	16	US-10-413-432-49	Sequence 49, Appl
23	456	15.9	574	14	US-10-032-233-40	Sequence 40, Appl
24	456	15.9	574	16	US-10-324-466-40	Sequence 40, Appl
25	456	15.9	574	16	US-10-413-432-40	Sequence 40, Appl
26	455	15.9	574	9	US-09-748-739A-8	Sequence 8, Appl
27	452	15.8	574	14	US-10-032-233-34	Sequence 34, Appl
28	452	15.8	574	16	US-10-324-466-34	Sequence 34, Appl
29	452	15.8	574	16	US-10-413-432-34	Sequence 34, Appl
30	452	15.8	614	15	US-10-116-275-258	Sequence 258, App
31	451	15.7	574	9	US-09-748-739A-6	Sequence 6, Appl
32	451	15.7	574	9	US-09-748-739A-17	Sequence 17, Appl
33	451	15.7	574	9	US-09-748-739A-20	Sequence 20, Appl
34	451	15.7	574	10	US-09-997-209-89	Sequence 89, Appl
35	451	15.7	574	14	US-10-032-233-22	Sequence 22, Appl
36	451	15.7	574	14	US-10-032-233-24	Sequence 24, Appl
37	451	15.7	574	14	US-10-032-233-26	Sequence 26, Appl
38	451	15.7	574	14	US-10-032-233-44	Sequence 44, Appl
39	451	15.7	574	14	US-10-032-233-47	Sequence 47, Appl
40	451	15.7	574	15	US-10-326-892-2	Sequence 2, Appl
41	451	15.7	574	16	US-10-433-206-89	Sequence 89, Appl
42	451	15.7	574	16	US-10-324-466-22	Sequence 22, Appl
43	451	15.7	574	16	US-10-324-466-24	Sequence 24, Appl
44	451	15.7	574	16	US-10-324-466-26	Sequence 26, Appl
45	451	15.7	574	16	US-10-324-466-44	Sequence 44, Appl

ALIGNMENTS

RESULT 1
US-09-943-857-4
; Sequence 4, Application US/09943857
; Publication No. US20030124701A1
; GENERAL INFORMATION:
; APPLICANT: Shaw, Jei-Fu
; APPLICANT: Lee, Guan-Chiun
; APPLICANT: Tang, Shye-Jye
; TITLE OF INVENTION: RECOMBINANT CANDIDA RUGOSA LIPASES
; FILE REFERENCE: 08919-066001
; CURRENT APPLICATION NUMBER: US/09/943,857
; CURRENT FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 547
; TYPE: PRT
; ORGANISM: Candida rugosa
; US-09-943-857-4

Query Match 100.0%; Score 2864; DB 10; Length 547;
Best Local Similarity 100.0%; Pred. No. 2.8e-265;
Matches 547; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	SMNSRGPAGRLGVSPTAKLANGTTITGLNAINAEFLGIPFAEPVGNLRFKDPVPYSGS	60
Db	1	SMNSRGPAGRLGVSPTAKLANGTTITGLNAINAEFLGIPFAEPVGNLRFKDPVPYSGS	60
Qy	61	LNQCKTSTYGPSQCNQPECTFEENLGKTDALVMSQKVFQAVLPOSEDCLTINVVRPPG	120
Db	61	LNQCKTSTYGPSQCNQPECTFEENLGKTDALVMSQKVFQAVLPOSEDCLTINVVRPPG	120
Qy	121	TKAGANLPVWLWIFGGGFEIGSTPIPPAOMVTKSVLMGKHIIHVAIVYVAVSWGFLAGD	180
Db	121	TKAGANLPVWLWIFGGGFEIGSTPIPPAOMVTKSVLMGKHIIHVAIVYVAVSWGFLAGD	180
Qy	181	DIKAEGSGNAGLKQDRGLGMQWADNIAGFGDPSKVTIFGESAGSMVLCHLIWNDCNT	240

Db 181 DIKAEAGNAGLKQORLQMGQWADNIAGFGDPSKVTFIGESAGSMSVLCHLIWNGDNT 240
Qy 241 YKGKPLFRAGIMQSGAMVPSDPVDTGYGNEIYDLFVSSAGCGSADKLACLRSSDITLL 300
Db 241 YKGKPLFRAGIMQSGAMVPSDPVDTGYGNEIYDLFVSSAGCGSADKLACLRSSDITLL 300
Qy 301 DATNNTPGFLAYSSLRSLSYLPRPDGKNTTDDMYKLVRDGKVASVPVVIIGDQNDGDTIFGL 360
Db 301 DATNNTPGFLAYSSLRSLSYLPRPDGKNTTDDMYKLVRDGKVASVPVVIIGDQNDGDTIFGL 360
Qy 361 SSLNVTNNAQARAYFKQSFIIHASDAEIDTLMAAYPQDITQGSPPDTGVNALTPOFKRIS 420
Db 361 SSLNVTNNAQARAYFKQSFIIHASDAEIDTLMAAYPQDITQGSPPDTGVNALTPOFKRIS 420
Qy 421 AVLDGLAFIHARRYFLNHFGQGTKYFSLSKOLSGLPIMGTFPHANDIWMQDYLILGSGSVIY 480
Db 421 AVLDGLAFIHARRYFLNHFGQGTKYFSLSKOLSGLPIMGTFPHANDIWMQDYLILGSGSVIY 480
Qy 481 NNAFIATDLPDNTAGLLVNWPKYTSSQSGNNLMINALGLYTGKDNFRTAGYDALMT 540
Db 481 NNAFIATDLPDNTAGLLVNWPKYTSSQSGNNLMINALGLYTGKDNFRTAGYDALMT 540
Qy 541 NPSSFFV 547
Db 541 NPSSFFV 547

RESULT 2
US-09-943-857-10
; Sequence 10, Application US/09943857
; Publication No. US20030124701A1
; GENERAL INFORMATION:
; APPLICANT: Shaw, Jai-Fu
; APPLICANT: Lee, Guan-Chiun
; APPLICANT: Tang, Shye-Jye
; TITLE OF INVENTION: RECOMBINANT CANDIDA RUGOSA LIPASES
; FILE REFERENCE: 08919-066001
; CURRENT APPLICATION NUMBER: US/09/943.857
; CURRENT FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 547
; TYPE: PRT
; ORGANISM: Candida rugosa
US-09-943-857-10

Query Match 90.4%; Score 2588; DB 10; Length 547;
Best Local Similarity 88.7%; Pred. No. 8.1e-239;
Matches 485; Conservative 30; Mismatches 32; Indels 0; Gaps 0;

Qy 1 SMNSRGPGAGRLGSVPTAKLANGDTITGLNAINAEFLGIPFAEPVGNLRKDPVPYSGS 60
Db 1 SMNSRGPGAGRLGSVPTATLANGDTITGLNAINAEFLGIPFAEPVGNLRKDPVPYSGS 60
Qy 61 LKGKFTSYGSCMQONPEGTFEENLKGKALDLMQSKVFOAVLPQSEDCLTINVRPPG 120
Db 61 LDGKFTSYGSCMQONPEGTFEENLPKAAALDLMQSKVFAVSPSEDCLTINVRPPG 120
Qy 121 TKAGANLPVMLWIFGGGFEIGSPFIAPPQWMTKSVLMGKHIHVAVNRYVASSGFLAGD 180
Db 121 TKAGANLPVMLWIFGGGFEVGGTSTFPQAWMTKSIAMGKPIIHVSNRYVSSGFLAGD 180
Qy 181 DIKAEAGNAGLKQORLQMGQWADNIAGFGDPSKVTFIGESAGSMSVLCHLIWNGDNT 240
Db 181 EIKAEAGNAGLKQORLQMGQWADNIAGFGDPSKVTFIGESAGSMSVWCHLIWNGDNT 240
Qy 241 YKGKPLFRAGIMQSGAMVPSDPVDTGYGNEIYDLFVSSAGCGSADKLACLRSSDITLL 300
Db 241 YKGKPLFRAGIMQSGAMVPSDAVDGVYGNEIFDILLASDAGCGSADKLACLRGVSSDTLE 300
Qy 301 DATNNTPGFLAYSSLRSLSYLPRPDGKNTTDDMYKLVRDGKVASVPVVIIGDQNDGDTIFGL 360

Db 301 DATNNTPGFLAYSSLRSLSYLPRPDGKNTTDDMFALVREGKVASVPVVIIGDQNDGDTFFGT 360
Qy 361 SSLNVTNNAQARAYFKQSFIIHASDAEIDTLMAAYPQDITQGSPPDTGVNALTPOFKRIS 420
Db 361 SSLNVTNNAQARAYFKQSFIIHASDAEIDTLMAAYPQDITQGSPPDTGVNALTPOFKRIS 420
Qy 421 AVLDGLAFIHARRYFLNHFGQGTKYFSLSKOLSGLPIMGTFPHANDIWMQDYLILGSGSVIY 480
Db 421 AVLDGLAFIHARRYFLNHFGQGTKYFSLSKOLSGLPIMGTFPHANDIWMQDYLILGSGSVIY 480
Qy 481 NNAFIATDLPDNTAGLLVNWPKYTSSQSGNNLMINALGLYTGKDNFRTAGYDALMT 540
Db 481 NNAFIATDLPDNTAGLLVNWPKYTSSQSGNNLMINALGLYTGKDNFRTAGYDALMT 540
Qy 541 NPSSFFV 547
Db 541 NPSSFFV 547

RESULT 3
US-09-943-857-8
; Sequence 8, Application US/09943857
; Publication No. US20030124701A1
; GENERAL INFORMATION:
; APPLICANT: Shaw, Jai-Fu
; APPLICANT: Lee, Guan-Chiun
; APPLICANT: Tang, Shye-Jye
; TITLE OF INVENTION: RECOMBINANT CANDIDA RUGOSA LIPASES
; FILE REFERENCE: 08919-066001
; CURRENT APPLICATION NUMBER: US/09/943.857
; CURRENT FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 547
; TYPE: PRT
; ORGANISM: Candida rugosa
US-09-943-857-8

Query Match 87.1%; Score 2495; DB 10; Length 547;
Best Local Similarity 86.3%; Pred. No. 6.7e-230;
Matches 472; Conservative 31; Mismatches 44; Indels 0; Gaps 0;

Qy 1 SMNSRGPGAGRLGSVPTAKLANGDTITGLNAINAEFLGIPFAEPVGNLRKDPVPYSGS 60
Db 1 SMNSRGPGAGRLGSVPTATLANGDTITGLNAINAEFLGIPFAEPVGNLRKDPVPYSGS 60
Qy 61 LKGKFTSYGSCMQONPEGTFEENLKGKALDLMQSKVFOAVLPQSEDCLTINVRPPG 120
Db 61 LKGKFTSYGSCMQONPEGTFEENLPKVALDLMQSKVFOAVLPQSEDCLTINVRPPG 120
Qy 121 TKAGANLPVMLWIFGGGFEIGSPFIAPPQWMTKSVLMGKHIHVAVNRYVASSGFLAGD 180
Db 121 TKAGANLPVMLWIFGGGFEIGSPFIAPPQWMTKSVLMGKHIHVAVNRYVASSGFLAGD 180
Qy 181 DIKAEAGNAGLKQORLQMGQWADNIAGFGDPSKVTFIGESAGSMSVLCHLIWNGDNT 240
Db 181 DIKAEAGNAGLKQORLQMGQWADNIAGFGDPSKVTFIGESAGSMSVLCHLIWNGDNT 240
Qy 241 YKGKPLFRAGIMQSGAMVPSDPVDTGYGNEIYDLFVSSAGCGSADKLACLRSSDITLL 300
Db 241 YKGKPLFRAGIMQSGAMVPSDPVDTGYGCAQIYDITLVASTGSSASNKACLRIGLSTQALL 300
Qy 301 DATNNTPGFLAYSSLRSLSYLPRPDGKNTTDDMYKLVRDGKVASVPVVIIGDQNDGDTIFGL 360
Db 301 DATNNTPGFLAYSSLRSLSYLPRPDGKNTTDDMYKLVRDGKVASVPVVIIGDQNDGDTIFGL 360
Qy 361 SSLNVTNNAQARAYFKQSFIIHASDAEIDTLMAAYPQDITQGSPPDTGVNALTPOFKRIS 420
Db 361 SSLNVTNNAQARAYFKQSFIIHASDAEIDTLMAAYPQDITQGSPPDTGVNALTPOFKRIS 420
Qy 421 AVLDGLAFIHARRYFLNHFGQGTKYFSLSKOLSGLPIMGTFPHANDIWMQDYLILGSGSVIY 480


```
Db 421 AVLGLDFTLTSRRYFLNHVYTGPKYFSLSKQSLGLPILGTTHANDIVWQHFLILGSGSVIY 480
QY 481 NNAFTAFATDLDPTAGLLVNMKPTSSSQSGNNLMINALGLYTGKDNFRTAGYDALMT 540
Db 481 NNAFTAFATDLDPTAGLLVNMKPTSSSQSGNNLMINALGLYTGKDNFRTAGYDALMT 540
QY 541 NPSSFFV 547
Db 541 DPSSFFV 547

RESULT 4
US-09-943-857-6
; Sequence 6, Application US/09943857
; Publication No. US20030124701A1
; GENERAL INFORMATION:
; APPLICANT: Shaw, Jui-Fu
; APPLICANT: Lee, Guan-Chiun
; TITLE OF INVENTION: RECOMBINANT CANDIDA RUGOSA LIPASES
; FILE REFERENCE: 08919-066001
; CURRENT APPLICATION NUMBER: US/09/943,857
; CURRENT FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 547
; TYPE: PRT
; ORGANISM: Candida rugosa
US-09-943-857-6

Query Match 85.0%; Score 2434; DB 10; Length 547;
Best Local Similarity 83.5%; Pred. No. 4.7e-224;
Matches 457; Conservative 38; Mismatches 52; Indels 0; Gaps 0;

QY 1 SMNSRGPAGRLGSVPTAKLANGDTITGLNAINAEFLGIPPAEPVGNLRFKDPVPYSGS 60
Db 1 SMNSRGPAGRLGSVPTAKLANGDTITGLNAINAEFLGIPPAEPVGNLRFKDPVPYSGS 60
QY 61 LINGQFTSYGSPCMQNPGEPTPEENLGTALDLVMSKVQAVLPQSEDECLINVRPPG 120
Db 61 LINGQFTSYGSPCMQNPGLNWDLSLPKAAINSIMQSKLFQAVLPNGEDECLINVRPPG 120
QY 121 TKAGANLPMWLIFGGGFEIGSPTIPPPAQMTKSVLMGKHIIHVAVNVRVASWGFLAGD 180
Db 121 TKPGANLPMWLIFGGGFEVGGSSLPFAQMITASVLMGKPIIHVSMNVRVASWGFLAGD 180
QY 181 DIKAGSGNAGLKDORLGMQWADNIAGFGDPKVTIFGESAGSMVLCGLIWNDDGNT 240
Db 181 DIKAGSGNAGLHQRLGLQWADNIAGFGDPKVTIFGESAGSMVLMCQLLWDDGNT 240
QY 241 YGKPLFRAGIMQSGAMVPSDVPDGTGNEIYDLFVSSAGCGSADKLACLSASDITLL 300
Db 241 YNGKPLFRAAIMQSGMVPSPDVPDGTGTEIYQVVASAGCGSADKLACLSASDITLL 300
QY 301 DATNTPGLAYSLRLSLRPDGNITDDMYKLVRDQKYASVPVILGQNDDEGTIFGL 360
Db 301 YGKPLFRAGIMQSGAMVPSDVPDGTGTEIYQVVASAGCGSADKLACLSASDITLL 360
QY 361 SSLNVTNQAARAYFKQSFHASDAEIDTLMAAYPQDITQGSPPDGTGVNALTPOPKRIS 420
Db 361 SSLNVTNQAARAYFKQSFHASDAEIDTLMAAYPQDITQGSPPDGTGVNALTPOPKRIS 420
QY 421 AVLGLDAFTHARRYFLNHVYTGPKYFSLSKQSLGLPILGTTHANDIVWQHFLILGSGSVIY 480
Db 421 AVLGLDAFTLPRYFLNHVYTGPKYFSLSKQSLGLPILGTTHANDIVWQHFLILGSGSVIY 480
QY 481 NNAFTAFATDLDPTAGLLVNMKPTSSSQSGNNLMINALGLYTGKDNFRTAGYDALMT 540
Db 481 NNAFTAFATDLDPTAGLLVNMKPTSSSQSGNNLMINALGLYTGKDNFRTAGYDALMT 540
QY 541 NPSSFFV 547
Db 541 DPSSFFV 547
```

```
RESULT 6
US-10-369-493-4236
; Sequence 4236, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
```

```
Db 541 NPSSFFV 547

RESULT 5
US-09-943-857-2
; Sequence 2, Application US/09943857
; Publication No. US20030124701A1
; GENERAL INFORMATION:
; APPLICANT: Shaw, Jui-Fu
; APPLICANT: Lee, Guan-Chiun
; TITLE OF INVENTION: RECOMBINANT CANDIDA RUGOSA LIPASES
; FILE REFERENCE: 08919-066001
; CURRENT APPLICATION NUMBER: US/09/943,857
; CURRENT FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 547
; TYPE: PRT
; ORGANISM: Candida rugosa
US-09-943-857-2

Query Match 83.7%; Score 2396; DB 10; Length 547;
Best Local Similarity 82.3%; Pred. No. 2.1e-220;
Matches 450; Conservative 34; Mismatches 63; Indels 0; Gaps 0;

QY 1 SMNSRGPAGRLGSVPTAKLANGDTITGLNAINAEFLGIPPAEPVGNLRFKDPVPYSGS 60
Db 1 SMNSRGPAGRLGSVPTAKLANGDTITGLNAINAEFLGIPPAEPVGNLRFKDPVPYSGS 60
QY 61 LINGQFTSYGSPCMQNPGEPTPEENLGTALDLVMSKVQAVLPQSEDECLINVRPPG 120
Db 61 LINGQFTSYGSPCMQNPWGSFEDTLPKNALDLVLSKLFQVVLNDEDECLINVRPPG 120
QY 121 TKAGANLPMWLIFGGGFEIGSPTIPPPAQMTKSVLMGKHIIHVAVNVRVASWGFLAGD 180
Db 121 TRASAGLPMWLIFGGGFELGSSLPFGQWAKSVLMGKPVIVHSMNVRVASWGFLAGD 180
QY 181 DIKAGSGNAGLKDORLGMQWADNIAGFGDPKVTIFGESAGSMVLCGLIWNDDGNT 240
Db 181 DIQNGSGNAGLHQRLGLQWADNIAGFGDPKVTIYGESAGSMSTFVHLVWDDGNT 240
QY 241 YGKPLFRAGIMQSGAMVPSDVPDGTGNEIYDLFVSSAGCGSADKLACLSASDITLL 300
Db 241 YNGKPLFRAAIMQSGMVPSPDVPDGTGTEIYQVVASAGCGSADKLACLSASDITLL 300
QY 301 DATNTPGLAYSLRLSLRPDGNITDDMYKLVRDQKYASVPVILGQNDDEGTIFGL 360
Db 301 QATSDTPGLAYSLRLSLRPDGTFTDDMYALVRDQKYAHVPVILGQNDDEGTIFGL 360
QY 361 SSLNVTNQAARAYFKQSFHASDAEIDTLMAAYPQDITQGSPPDGTGVNALTPOPKRIS 420
Db 361 SSLNVTNQAARAYFKQSFHASDAEIDTLMAAYTSDITQGSPPDGTGVNALTPOPKRIS 420
QY 421 AVLGLDAFTHARRYFLNHVYTGPKYFSLSKQSLGLPILGTTHANDIVWQHFLILGSGSVIY 480
Db 421 ALLGLDAFTLARRYFLNHYVYQGTQYFSLSKQSLGLPVLGTGHGNDIIMQDYLVGSGSVIY 480
QY 481 NNAFTAFATDLDPTAGLLVNMKPTSSSQSGNNLMINALGLYTGKDNFRTAGYDALMT 540
Db 481 NNAFTAFANDLDPNKAGLWNTWPTYTSSSQSGNNLMINGLGLYTGKDNFRPRDAYSAF 540
QY 541 NPSSFFV 547
Db 541 NPSSFFV 547
```

```

; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 4236
; LENGTH: 561
; TYPE: PRT
; ORGANISM: Neurospora crassa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(561)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-4236

Query Match 40.9%; Score 1170.5; DB 15; Length 561;
Best Local Similarity 43.6%; Pred. No. 7.2e-103;
Matches 247; Conservative 96; Mismatches 175; Indels 49; Gaps 10;

Qy 13 SVPTAKLANGDTITGLNAINEAEFLGIPPAEPVGNLRFKDPVPPYSGSLNGQKTSYGPS 72
Db 1 AAPTPTITSTG-TIVGANGILTEAFNGIPYALPPTGNLRLKPPVRLKSSLGVSDFDASGIGPA 59
Qy 73 CMOQNPGTEFEENLGKLTALDVMQSKVFOAVLPQSEDCLTINVVRPPTKAGANLPVLMW 132
Db 60 CPQFLATSSNEFLPQ-VIDKIVNTQFLKILNVKEDCLISVTRPKTKAGDKLPVLF 118
Qy 133 IFGGGFPI-----GSPTIFPPAQMTKSVLMGKHIIHVAVNRYVASW 174
Db 119 IFGGGFVVRNPAYTRDVMNCKNFTAGSASMYDGAFLVTNAINMGKPYVYVAVNRYVGGF 178
Qy 175 GFLAGDDIKAGSGNAGLKQORLCQWQVADNIAGFGGDPKVTIFGSGAGSMVLCHLIW 234
Db 179 GFMPGKBLKDGSSNLGHLDDQRMGLQWVADNIAAFGGDDPKVTIWSGAGSMVFNQMSL 238
Qy 235 NGDNTYKGLPLFRAGIMQSGAMVPSDPVDTGTYGNEIYDLFVSSAGSGSADKLCLRSA 294
Db 239 YDGNNTYNGKPLFRGAIMWSSGIYPAGVDCPKQKYDITVKNAGSGSADTLCLRAL 298
Qy 295 SSDTLLDATNTPGFLAYSSRLSYLPRPGKNTDDMYKLVRDGKVASVPVIGDQND 354
Db 299 PYETFLKAANSVPGLSYNSVALSYLPRPGKALTQSADKLMILAKKYAAVPMIIGDQDE 358
Qy 355 GTIFGLSLNVTNAQARAYFKQSPFIH-ASDAEIDTLMAAYPODITOGSPDTCVLNALT 413
Db 359 GTLPSLFQSNITTSKLVSYNDLFFNDATSEQIKLSVYSTLTISAGSPGGLFNEIY 418
Qy 414 PQFKRISAVLGLDLAFIHARRYFLNHFQGGT-----KYSFLSKQLSGPLMTGTFHANDI 466
Db 419 PGFKRLAAILGLDILFTLSRIIFLD--AATLNPSPVAMSVLASYNFTGPIGLTFFHASDI 475
Qy 467 --VWQDYLLGSGSVIYNNAIFATDLDPN--TAGLLVNWPKYTSSSQSGNNLM 516
Db 476 LOVFYGLPLNYPASKSIQSYVANFYNLDPNNDASGTSKSKSVQDMPWQOKERK----- 529
Qy 517 MINALGYLTG--KDNRTAGYDALMTN 541
Db 530 LVQFFSDYAGVLADDFRSDSYNWKAN 556

```

RESULT 7
 US-09-875-353-4
 ; Sequence 4, Application US/09875353
 ; Patent No. US20020168713A1
 ; GENERAL INFORMATION:

```

; APPLICANT: Curtis, Rory A.J.
; TITLE OF INVENTION: 46980, A NOVEL HUMAN NEUROLIGIN FAMILY
; TITLE OF INVENTION: MEMBER AND USES THEREOF
; FILE REFERENCE: 10448-058001
; CURRENT APPLICATION NUMBER: US/09/875,353
; CURRENT FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/209,949
; PRIOR FILING DATE: 2000-06-06
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 612
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: consensus sequence
US-09-875-353-4

Query Match 23.6%; Score 674.5; DB 9; Length 612;
Best Local Similarity 34.0%; Pred. No. 3e-55;
Matches 199; Conservative 80; Mismatches 203; Indels 103; Gaps 19;

Qy 19 LANGDITITGLNAINE-----AFLGIPPAEPVGNLRFKDPVPPY-----SGSLNGQ 64
Db 31 VATNNVLGKVRGVNTEKTDNGEQSVSYFLGIPYAEPPVGNLRFKAPQFYKEPNSDVLDA- 89
Qy 65 KFTSYGSCMQQPEGTFFENLGKLTALDVMQSKVFOAVLPQSEDCLTINVVRPPTKAG 124
Db 90 --TKYPSCQLQDDDFGFSLSDL-KVALKMLSLGNKLVGLKLSDECLYLVNYPKTKPN 146
Qy 125 ANLPVLMWIFGCGFEIGSPITFPPAQMTKSVLMGKHIIHVAVNRYVASWGLAGDDIKA 184
Db 147 SKLPVVMVHGGGFMFGSGHSLPLSYDGSGLAREGNVIVVSINRYRLGFLGLSTGDDKL 206
Qy 185 EGSNAGLKQORLCQWQVADNIAGFGGDPKVTIFGSGAGSMVLCHLIWNDCGNTYK 244
Db 207 PGSGNYGLLDORLALKWVQDNIAAFGGDPNSVTIFGSGAGASVSLLLSNGGDNPPSSK 266
Qy 245 PLFRAGIMQSGA-----MVPSPDVGTYGNEIYDLFVSSAGCG--SASDKLACLRSASSD 297
Db 267 GLFHRAISQSGSALSFWAIQSESNARGAKELARLL-----GCNETSSELDCLRSAE 322
Qy 298 TLLDATNTPGFLAYS-----SLRLSYLPRPGKN-----ITDDMYKLVRDGKVASVPVI 347
Db 323 ELLEATRS---FLLEFVVPFLPLFLAPGVVDGDAPEAFIPEDPELIKEGKPADVPYL 379
Qy 348 IGOQNDGCTIFGLSSLNVTNAQAR-----AYPKQSFHASDA-----EID 388
Db 380 IGVTKDEGGYFAAMLNASKGDELDKKTNPDPVWLELLKYLIFYASEALNIKMDMLAD 439
Qy 389 TLMAAYPODITOGSPDTCVLNALTPOFKRISAVLGLDLAF-----THARR----- 433
Db 440 KVLKPYDGVDDFS-----VESRKNLQDMLTDLKPCPTRVAADLHAKHGSGSPVY 490
Qy 434 -YFLNHFQGGTKSYFLSKQLSGPLMTGTFHANDI--VWQDYLL-----GSGSV 478
Db 491 AYVFDHPASFGIGQFLAKRVDP-EFGAVHGDIEFFVFGNPLLKEQLYKATEBEKSSSK 549
Qy 479 IYNNAIFATDLDPN--TAGLLVNWPKYTSSSQSGNNLMINAL 521
Db 550 TMNMYANFAKTGNPNNGTNSGLVWVWPKYTSSEQKYSLLILLTTI 594

```

RESULT 8
 US-10-023-515-4
 ; Sequence 4, Application US/10023515
 ; Publication No. US20020182636A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Curtis, Rory A. J.
 ; APPLICANT: Silos-Santiago, Immaculada
 ; TITLE OF INVENTION: 53010, A NOVEL HUMAN CARBOXYLESTERASE
 ; TITLE OF INVENTION: FAMILY MEMBER AND USES THEREOF
 ; FILE REFERENCE: 10448-122001

```
; CURRENT APPLICATION NUMBER: US/10/023,515
; CURRENT FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 60/256,369
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 60/279,508
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 574
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus sequence
US-10-023-515-4

Query Match      23.5%; Score 672.5; DB 13; Length 574;
Best Local Similarity 34.9%; Pred. No. 4.2e-55;
Matches 195; Conservative 76; Mismatches 195; Indels 93; Gaps 18;

QY 35 AFLGIPFAAPPVGNLRFKDPVPY----SGSLGQKFTSYGSCMQONPEGTPEENLGKTA 90
Db 19 SFLGIPYAEPPVGNLRFKAPQYKEPWSVDLDA---TKYPPSCLDQDDDFGFSLSDL-KVA 74

QY 91 LDLVMQSKVFQAVLPQSEDCLTINVRPPTKAGANLPVLMWIFGGGFEIGSTIIPPPAQ 150
Db 75 LKWLGLGNKLVGLKLSLSEDCLYLVNVTYKNTKPNKLPVWVWTHGGGFMFGSGHSLPLSL 134

QY 151 MYTKSVLMGKHIIHVAVNRYVASWGLAGDDDKAEGSGNAGLKQDRLGMQWADNITAGFG 210
Db 135 YDGESLAREGNVIVSYNRYLGLPLSTGDDKLPKSGNYGLLDQRLALKWQVDNIAAFG 194

QY 211 GPSPKVTTFEGSAGMSVLCHLIWNGDNTYKGLPFRAGIMOSGA----MVPSDPVDG 265
Db 195 GPNSTVTFEGSAGASVSLILLNSGDNPPSKGLFHRAISOGSALSFPWAIQSESNAR 254

QY 266 TYGNEYIDLVFSAGCG--SASDKLACLSASDITLLDATNNTPGPLAYS-----SLRLS 318
Db 255 GRAKELARLL---GCNETSSSELDDCLRSKSAEELLEATRS---FLLFEYVPFLPLFLA 307

QY 319 YLPRPDGKN-----ITDDMYKLVRDGKYASVPVIIGDQNDGTTIFGLSSINVTNNAQAR- 372
Db 308 FGPVVDGDDAPEAFIPEDPEELIKEGFADVPYLIIGVTKDEGGYFAAMLINASSKGEDEL 367

QY 373 -----AVFKQSFIHASDA-----EDTLMAAYPODITOGSPEDTGVNALTP 414
Db 368 KKETNPDWLELLKYLIFYASEALNKKMDLADKLVLEKYPGDVDDFS-----VES 418

QY 415 QPKRISAVLGLDAP-----IHARR-----YFLNHFGGTYKSYFLSKQLSGLPIM 458
Db 419 RKPNIQDMLTDLFLKCPTRVAADLHAKHGSPVYAVFDHPASFGIGQFLAKRVDP-EFG 477

QY 459 GFTHANDI--VMQDYLL-----GSGSVIYNNAFIATDLDN--TAGLLVNW 502
Db 478 GAVHGDEIFVFGNPLKQLYKATEEBEKSSKTMWNYANFAKTGNPNNGTNSGLVWV 537

QY 503 PKYTSSSQGNLMINAL 521
Db 538 PKYTSEEQYSLILLITTI 556

RESULT 9
US-10-674-636-4
; Sequence 4, Application US/10674636
; Publication No. US20040086922A1
; GENERAL INFORMATION:
; APPLICANT: Curtie, Rory A. J.
; TITLE OF INVENTION: 53010, A NOVEL HUMAN CARBOXYLSTERASE
; FILE REFERENCE: 10448-122001
; CURRENT APPLICATION NUMBER: US/10/674, 636
; CURRENT FILING DATE: 2003-09-29

; CURRENT APPLICATION NUMBER: US/10/023,515
; CURRENT FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 60/256,369
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 60/279,508
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 574
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus sequence
US-10-023-515-4

Query Match      23.5%; Score 672.5; DB 13; Length 574;
Best Local Similarity 34.9%; Pred. No. 4.2e-55;
Matches 195; Conservative 76; Mismatches 195; Indels 93; Gaps 18;

QY 35 AFLGIPFAAPPVGNLRFKDPVPY----SGSLGQKFTSYGSCMQONPEGTPEENLGKTA 90
Db 19 SFLGIPYAEPPVGNLRFKAPQYKEPWSVDLDA---TKYPPSCLDQDDDFGFSLSDL-KVA 74

QY 91 LDLVMQSKVFQAVLPQSEDCLTINVRPPTKAGANLPVLMWIFGGGFEIGSTIIPPPAQ 150
Db 75 LKWLGLGNKLVGLKLSLSEDCLYLVNVTYKNTKPNKLPVWVWTHGGGFMFGSGHSLPLSL 134

QY 151 MYTKSVLMGKHIIHVAVNRYVASWGLAGDDDKAEGSGNAGLKQDRLGMQWADNITAGFG 210
Db 135 YDGESLAREGNVIVSYNRYLGLPLSTGDDKLPKSGNYGLLDQRLALKWQVDNIAAFG 194

QY 211 GPSPKVTTFEGSAGMSVLCHLIWNGDNTYKGLPFRAGIMOSGA----MVPSDPVDG 265
Db 195 GPNSTVTFEGSAGASVSLILLNSGDNPPSKGLFHRAISOGSALSFPWAIQSESNAR 254

QY 266 TYGNEYIDLVFSAGCG--SASDKLACLSASDITLLDATNNTPGPLAYS-----SLRLS 318
Db 255 GRAKELARLL---GCNETSSSELDDCLRSKSAEELLEATRS---FLLFEYVPFLPLFLA 307

QY 319 YLPRPDGKN-----ITDDMYKLVRDGKYASVPVIIGDQNDGTTIFGLSSINVTNNAQAR- 372
Db 308 FGPVVDGDDAPEAFIPEDPEELIKEGFADVPYLIIGVTKDEGGYFAAMLINASSKGEDEL 367

QY 373 -----AVFKQSFIHASDA-----EDTLMAAYPODITOGSPEDTGVNALTP 414
Db 368 KKETNPDWLELLKYLIFYASEALNKKMDLADKLVLEKYPGDVDDFS-----VES 418

QY 415 QPKRISAVLGLDAP-----IHARR-----YFLNHFGGTYKSYFLSKQLSGLPIM 458
Db 419 RKPNIQDMLTDLFLKCPTRVAADLHAKHGSPVYAVFDHPASFGIGQFLAKRVDP-EFG 477

QY 459 GFTHANDI--VMQDYLL-----GSGSVIYNNAFIATDLDN--TAGLLVNW 502
Db 478 GAVHGDEIFVFGNPLKQLYKATEEBEKSSKTMWNYANFAKTGNPNNGTNSGLVWV 537

QY 503 PKYTSSSQGNLMINAL 521
Db 538 PKYTSEEQYSLILLITTI 556

RESULT 10
US-09-934-323-4
; Sequence 4, Application US/09934323
; Patent No. US20020150910A1
; GENERAL INFORMATION:
; APPLICANT: Curtie, Rory A. J.
; TITLE OF INVENTION: 33410, A NOVEL HUMAN CARBOXYLSTERASE
; FILE REFERENCE: 10448-081001
; CURRENT APPLICATION NUMBER: US/09/934,323
; CURRENT FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 60/226,774
```

```
; PRIOR FILING DATE: 2000-08-21
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 585
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus sequence
; US-09-934-323-4

Query Match      23.2%; Score 664; DB 9; Length 585;
Best Local Similarity 34.0%; Pred. No. 2.8e-54;
Matches 199; Conservative 80; Mismatches 203; Indels 104; Gaps 20;

Qy 19 LANGDTITGLNAIN-----AFLGIPFAEPVGNLRFKDPVPY-----SGSLNGQ 64
Db 3 VATNNVLGKVRGNEKTDNGEQSVYSLGIPYAEPPVGNLRFKAPQYKEPWSVDVLD- 61
Qy 65 KFTSYGSCMOONPEGTFEENLGTALDLVMSKVQAVLPQSEDCLTINVRPFGTKAG 124
Db 62 --TKYPPSCLODDFGSLSD- KVALKMLSLGNKLVGLKLSDCLYLVNVTYKYNKP 118
Qy 125 ANLPVLMWIFGGFEIGSPITFFPAQMVTKSVLMKGIHIVAVNRYVASWGLAGDDIKA 184
Db 119 SKLPVWIIHGGFMFGSGHSLPLSLYDGESLAREGNVIVVSINVRGLPLGFLSTGDDKL 178
Qy 185 EGSGNAG-LKDQRLGMQWADNIAGFGDPKSVTIFGESAGSMVSLCHLIWNGDNTYKG 243
Db 179 PGSGNYGLLLQRLALKWVQDNIAAFGGDPNSVTIFGESAGAAVSULLLSNGGDNPPSS 238
Qy 244 KPLFRAGIMQSGA----MVPSPDPVDTGYGNEIYDLFVSSAGCG--SASDKLACLRSASS 296
Db 239 KGLFRAISQSGSALSPWAIQSESNARGRAKELARLL-----GCNETSSELDDCLRSKA 294
Qy 297 DTLDDATNTPGFLAYS-----SLRLSYLPRPDGKN-----ITDDMYKLVRDGYKASVPV 346
Db 295 EELLEATRS---FLLFYVVPFLPLFLAPGVVDGDAPEAFIPEDPEBLIKEGKFADVPY 351
Qy 347 IIGQNDGTTPLGLSSLVNVTNAQAR-----AYFKQSFHASDA-----EI 387
Db 352 LIGVTKDEGGYFAAMLNASSGDELLAKETNPVWLELLKYLIFYASEALNIKMDMDLA 411
Qy 388 DTLMAAYPDITQSPDPTGVNLALTPOFKRISAVLGDLAF-----IHARR----- 433
Db 412 DKVLEKTPGVDDFS-----VESRKPMLQMDLTDLKPCPTRVAADLHAKHGSPV 462
Qy 434 --YFLNHFQGGTKYSFLSKQLSGLPIMGTFFHANDI--VWQDYLL-----GSGS 477
Db 463 YAYVFDHPASFGIGQLAKRVDP-EFGGAVHGDEIFFVFGNPLLKEQLYKATEBEEKSSS 521
Qy 478 VIYNNAFTAFATDLPN--TAGLLVWMPKYTSSQSGNNLMMINAL 521
Db 522 KTMNYWANFAKTGNPNNGTNGLVVWMPKYTSEBQKSYLLILLTTI 567

RESULT 11
US-09-748-739A-21
; Sequence 21, Application US/09748739A
; Patent No. US20020119489A1
; GENERAL INFORMATION:
; APPLICANT: Lockridge, Jeffrey D.
; APPLICANT: Watkins, Oksana
; TITLE OF INVENTION: Butyrylcholinesterase Variants and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: P-IX 4143
; CURRENT APPLICATION NUMBER: US/09/748,739A
; CURRENT FILING DATE: 2000-12-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 574
; TYPE: PRT

; ORGANISM: Equus caballus
US-09-748-739A-21
Query Match      16.8%; Score 482; DB 9; Length 574;
Best Local Similarity 30.4%; Pred. No. 7.8e-37;
Matches 162; Conservative 78; Mismatches 193; Indels 100; Gaps 23;

Qy 22 GDTITGLNAINAEFLGIPFAEPVGNLRFKDPVPYSGSLNGQKFTSYGSPSCMOONPEGT 81
Db 22 GGTVT-----AFLGIPYAPQPLGRLFKPKQSLTKWSINWATKYANSC-YQNTDQS 72
Qy 82 FEENLGTALDLVMSKVQAVLPQSEDCLTINVRP-PGTKAGANLPVMLWIFGGGPEI 140
Db 73 FPGFLG-----SEWNPNTSELSDCLYLVNWIAPKPK---NATVMIWYGGSPQT 120
Qy 141 GSPITFPPAQMVTKSVLMGK-----HIHIVAVNRYVASWGLAGDDIKABSGSNAGLKD 194
Db 121 GTSSL-----PVYDGKFLARVERVI VVSMNRYVGALGFLALSE-NPEAPGNMGLFD 170
Qy 195 QRLGMQWADNIAGFGDPKSVTIFGESAGSMVSLCHLIWNGDNTYKGKPLFRAGIMQS 254
Db 171 QQLALQWVQKNIAAFGCGNPRSVTLFESAGAAVSLSHL-----SPRSQPLFTRAILQS 224
Qy 255 GAMVSPDPVDTGYGNEIYDLFVSS-AGCG--SASDKLACLRSASSDTLLDNTNTPGFLA 311
Db 225 GSSNAPWAVTSLYEARNRTLAKRMGCSRDNETEMIKCLDKDQPEIL---LNEVFVVP 281
Qy 312 YSS-LRLSYLPRPDGKNITDDMYKLVRDGYKASVPVLIQNDRGTF-----GLSSLN 364
Db 282 YDTLLSVNFGTVDGDELTDPDTLLQLQGFQKRTQILVNVNKDEGTAFVYGAFGSKDN 341
Qy 365 --VTTNAQARAYFKQSFHASDABIDTLMAAY-----PQDITQSPDPTGVNLALT 414
Db 342 NSIITRKEFQGLKIPFPRVSEFGRESILFHYMDWLDQRAENVREALDDVVGVDYNIICP 401
Qy 415 --QPKRISAVLGDLAFTHARRYFLNHFQGGTKYSFLSKQLSGLPIMGTFFHANDI 472
Db 402 ALEFTRKFSGLNDADF---YYPEH-----RSTKLPWPPEMGMVHGYEI---EFV 445
Qy 473 LG---SGSVYNNNA-----FIATATDLPN--TAGLLVWMPKYTSSSQ 510
Db 446 FGLPLERRVNYTRABEILSRKIMKRWANFAKGNPNGTQNNSTRWPVFKSTEQ 498

RESULT 12
US-10-032-233-48
; Sequence 48, Application US/10032233
; Publication No. US20030153062A1
; GENERAL INFORMATION:
; APPLICANT: Watkins, Jeffrey D.
; APPLICANT: Hancock, James D.
; TITLE OF INVENTION: Butyrylcholinesterase Variants with
; TITLE OF INVENTION: Increased Catalytic Efficiency and Methods of Use
; FILE REFERENCE: P-IX 4642
; CURRENT APPLICATION NUMBER: US/10/032,233
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 574
; TYPE: PRT
; ORGANISM: Equus caballus
US-10-032-233-48

Query Match      16.8%; Score 482; DB 14; Length 574;
Best Local Similarity 30.4%; Pred. No. 7.8e-37;
Matches 162; Conservative 78; Mismatches 193; Indels 100; Gaps 23;

Qy 22 GDTITGLNAINAEFLGIPFAEPVGNLRFKDPVPYSGSLNGQKFTSYGSPSCMOONPEGT 81
Db 22 GGTVT-----AFLGIPYAPQPLGRLFKPKQSLTKWSINWATKYANSC-YQNTDQS 72
Qy 82 FEENLGTALDLVMSKVQAVLPQSEDCLTINVRP-PGTKAGANLPVMLWIFGGGPEI 140
Db 22 GGTVT-----AFLGIPYAPQPLGRLFKPKQSLTKWSINWATKYANSC-YQNTDQS 72
Qy 82 FEENLGTALDLVMSKVQAVLPQSEDCLTINVRP-PGTKAGANLPVMLWIFGGGPEI 140
```



```

22 QY 22 GDTITGLNAINEARFLIGIPAEPPVGNMLRPKDPVPYSSGLNGQKFTSYGSCMOONPEGT 81
Db 22 GGTVT-----AFLGIPAAQPLGSLRFFKKPQPLANKWPDPVWATKYANSC-YONIDQA 72
QY 82 FEENLKGKTALDVLVMQSVFOAVLFPQSEDCLTINTVVRP-PGTKAGANLPVMLWIFGGGF 140
Db 73 FPGQG-----SEMNENTNLSEDCLVNLWIPVPRFK-----NATVMVWVYGGGFQT 120
QY 141 GSPTIFPPAQMTKSVLMGK-----HIHVAVNYRVASMGFLAGDDIKABGSGNAGLKD 194
Db 121 GTSSL-----PVYDGKFLTRVERVIVVSMNYRVGALGFLAPPG-NSEAPGNMGLFD 170
QY 195 QRLCMQWVADNIIAGFGDPSKVTTIFGSGAGSMVLCHLIWNDGNTYKGKPLFRAGIMQS 254
Db 171 QQLALQWIORNIAAFGNGPKSVTLFGSAGAAVSLSHL-----CPQSYPLFTRAIL 224
QY 255 GAMVPSPDPVGTGYNEIYDLFVSA---GCGSASDK--LACLRSSADSTLLDATNNTPGF 309
Db 225 GS--SNAPWAKHPPEEARNTLTLLAKFTIGCSKENEKEIITCLRSKQPOEIL---LNEKLV 279
QY 310 LAYSLSR-LSVLPRPDGKNIITDDMKVLDRCKYASVPVIGQDQDEGTF-----GLSS 362
Db 280 LPSDSIRSNPGPTVGDGFLDMPHETLLOLGRVYTAQLVGVNKDEGTAFIVYGAPFSK 339
QY 363 LN--VTNQAQARAYFKOSFIHASDAEIDTLLMAAYPQDITQSP-----FD--TGVINAL 412
Db 340 DNDSILITRREFQEGNLNMYFPGVSSLGKAEAILFYVVDWLDGQTPVEVYREAFDDIIGDYNII 399
QY 413 TP--QPKRISAVLGDLAFIHARRVFLNHFQGTKYVFLSKQLSGLPMTGTFEHANDIYWD 470
Db 400 CPALEFTKFAELINAF-----YFER-----RSSKLPPEWGMVWHGYEI---443
QY 471 YLLG---SGSVYINNAFIATADL-----DPN-TAGLLVNWPKYTSQQ 510
Db 444 FVFGFLPLRRVNYTRAEBIFSRSLMKTWANFAKYCHPNGTGNSVWVPVFTSTEQ 498

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 10:48:09 ; Search time 19 Seconds
(without alignments)
1486.285 Million cell updates/sec

Title: US-09-943-857-4

Perfect score: 2864

Sequence: 1 SMNSRGPAGRLGSLVPTAKLA.....DNFRTAGYDALMTNPSFFV 547

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/prodata/2/iaa/5A COMB.pep.*

2: /cgn2_6/prodata/2/iaa/5B COMB.pep.*

3: /cgn2_6/prodata/2/iaa/6A COMB.pep.*

4: /cgn2_6/prodata/2/iaa/6B COMB.pep.*

5: /cgn2_6/prodata/2/iaa/PCTUS COMB.pep.*

6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1107	38.7	563	3	US-08-362-525-12
2	672.5	23.5	574	4	US-10-023-515-4
3	491.5	17.2	575	1	US-08-348-920-1
4	488.5	17.1	575	1	US-08-348-920-2
5	454	15.9	614	3	US-08-446-100-25
6	452	15.8	600	2	US-08-370-156-4
7	452	15.8	600	3	US-08-814-095-4
8	452	15.8	600	3	US-08-975-084-1
9	452	15.8	602	6	5215909-11
10	452	15.8	614	1	US-07-732-962A-2
11	452	15.8	614	2	US-08-370-156-2
12	452	15.8	614	3	US-08-446-100-19
13	452	15.8	614	3	US-08-446-100-21
14	452	15.8	614	3	US-08-814-095-2
15	452	15.8	614	5	PCT-US92-06106-2
16	452	15.8	617	3	US-08-370-156-6
17	452	15.8	617	3	US-08-814-095-6
18	451	15.7	602	3	US-08-446-100-1
19	451	15.7	602	3	US-08-446-100-24
20	451	15.7	602	3	US-09-334-489-3
21	451	15.7	602	3	US-09-334-489-4
22	449.5	15.7	572	2	US-08-932-376A-2
23	449	15.7	602	3	US-08-446-100-13
24	448	15.6	602	3	US-08-446-100-11
25	447	15.6	602	3	US-08-446-100-3
26	447	15.6	602	3	US-08-446-100-4
27	447	15.6	602	3	US-08-446-100-5

28 447 15.6 602 3 US-08-446-100-12 Sequence 12, Appl
29 447 15.6 602 3 US-08-446-100-15 Sequence 15, Appl
30 447 15.6 602 3 US-08-446-100-16 Sequence 16, Appl
31 447 15.6 602 3 US-08-446-100-17 Sequence 17, Appl
32 447 15.6 602 3 US-08-446-100-18 Sequence 18, Appl
33 447 15.6 614 3 US-08-446-100-22 Sequence 22, Appl
34 447 15.6 614 3 US-08-446-100-23 Sequence 23, Appl
35 446 15.6 602 3 US-08-446-100-8 Sequence 8, Appl
36 445 15.5 602 3 US-08-446-100-14 Sequence 14, Appl
37 444.5 15.5 544 2 US-08-932-376A-4 Sequence 4, Appl
38 444.5 15.5 635 6 5215909-10 Patent No. 5215909
39 444 15.5 614 3 US-08-446-100-20 Sequence 20, Appl
40 443 15.5 573 6 5215909-12 Patent No. 5215909
41 443 15.5 602 3 US-08-446-100-2 Sequence 2, Appl
42 443 15.5 602 3 US-08-446-100-6 Sequence 6, Appl
43 442 15.4 602 3 US-08-446-100-7 Sequence 7, Appl
44 442 15.4 602 3 US-08-446-100-9 Sequence 9, Appl
45 442 15.4 602 3 US-08-446-100-10 Sequence 10, Appl

ALIGNMENTS

RESULT 1

US-08-362-525-12
; Sequence 12, Application US/08362525
; Patent No. 6027910

; GENERAL INFORMATION:

; APPLICANT: KLIS, FRANCISCUS M.

; APPLICANT: SCHREUDER, MAARTEN P.

; APPLICANT: TOSCHKA, HOLSER Y.

; APPLICANT: VERRIPS, CORNELIS T.

; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE

; TITLE OF INVENTION: CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION

; TITLE OF INVENTION: PROTEIN

; NUMBER OF SEQUENCES: 26

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.

; STREET: 1100 New York Avenue, N.W.

; CITY: Washington

; STATE: D. C.

; COUNTRY: U.S.A.

; ZIP: 20005-3918

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/362,525

; FILING DATE: 04-JAN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: EP 92202080.5

; FILING DATE: 08-JUL-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: EP 92203899.7

; FILING DATE: 14-DEC-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/EP93/01763

; FILING DATE: 07-JUL-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: KOKULIS, PAUL N.

; REGISTRATION NUMBER: 16,773

; REFERENCE/DOCKET NUMBER: 213289/T7020 (V)

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 861-3000

; TELEFAX: (202) 822-0944

; TELEX: 6714627 CUSH

; INFORMATION FOR SEQ ID NO: 12:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 563 amino acids

; TYPE: amino acid


```
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-362-525-12

Query Match      38.7%; Score 1107; DB 3; Length 563;
Best Local Similarity 43.4%; Pred. No. 5.4e-100;
Matches 240; Conservative 73; Mismatches 196; Indels 44; Gaps 9;

Qy 9  GRLGSVTAKLANGDTITGLNAINAEPLGTIPFAEPVGNLRFPKDPYPSGSLNGQKETS 68
Db 16 GTLAQAATAVLNGNEVSGLEGKVDTPKGFIPFADPPVGDURFKHPQPFCTSGQLKAND 75
Qy 69 YGPSQMOQNPETGE-----ENLGKTAIDLVMQSKVFQAVLPQSEDCLTINV 115
Db 76 FSSACQLDPGNASLLDKVVLGKILPDNLGRPLYDMA-----QGSVSNEDCLYINV 129
Qy 116 VRPPTKAGANLPVLMITFGGGFEIGSTPIPPAQMTKSVLMGKHIIHVAVNYRVASWG 175
Db 130 FRPAGTRKPAKLPVMWIIYGAFAVFGSSASYPGNGYVKESVEMGQPVVVFVSINRYRTGPG 189
Qy 176 FLAGDDIKAESGNAGLKQRLGMQWADNIAGGDPDSKYTIIFGESAGMSVLCILWN 235
Db 190 FLGGDAITAEGNTNAGLHDQRKLEWSDNIANFGDDPKYMIIFGESAGAMVAHQLVAY 249
Qy 236 DGDNTYKGPPLFRAGIMOSGAMVPSDPVDGTGYGNEIYDLFVSSAGC---GSASDKLACL 292
Db 250 GGDNTYNGKQLFHSNAILQSGGLPYFDPSTSVGPESAYSRAFAQAGCDTSASDNDTLACL 309
Qy 293 SASSTLLDATNN-----TPGFLAYSRLSYLPRPGKNITDDMYKLVRDGYAS 343
Db 310 SKSSDVLHSAQNSYDLKDLFGLLPQLGFG-----PRPDGNIIPDAAYELYSRGYAK 362
Qy 344 VPVIIGQNDRGTTIFGLSSLVNTNAQARAYFKQSFHASDAEIDTLMAYPQDITGSP 403
Db 363 VPYITGNOEDEGTILAPVAINATTPHVKKLWKLYICSQASDASDLRDLVLSLTPGWSGSP 422
Qy 404 PDTGVNALTQPKRISAVLGDLAFIHARRYFLNHFQGTGKYSLSKOLSLG-PIMGTFH 462
Db 423 FRTGLNALTQPKRILAIPTDLLQSPRRVWLNATKDVNRWYLATQHLNVPLGTFH 482
Qy 463 ANDIVQDYL-LGSGSVIYNNAFIATFDLPNTAGLLVNVKPYTSSQSGNNLMINAL 521
Db 483 GSDLLFYQYVDLGPSSA-YRRYFISFANHDPNVGNTLQQWDMYT---DAGKEMLIQHMI 538
Qy 522 GLYTCKNFRTAG 534
Db 539 GNSMRTDDFRLEG 551

RESULT 2
US-10-023-515-4
; Sequence 4, Application US/10023515
; Patent No. 6664091
; GENERAL INFORMATION:
; APPLICANT: Curtis, Rory A. J.
; APPLICANT: Silos-Santiago, Inmaculada
; TITLE OF INVENTION: 53010. A NOVEL HUMAN CARBOXYLESTERASE
; TITLE OF INVENTION: FAMILY MEMBER AND USES THEREOF
; FILE REFERENCE: 10448-122001
; CURRENT APPLICATION NUMBER: US/10/023,515
; CURRENT FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 60/256,369
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 60/279,508
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 574
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus sequence

Query Match      23.5%; Score 672.5; DB 4; Length 574;
Best Local Similarity 34.9%; Pred. No. 3.3e-57;
Matches 195; Conservative 76; Mismatches 195; Indels 93; Gaps 18;

Qy 35 AFLGIPFAEPVGNLRFKDPVY-----SGSLNGQKFTSYGSCMQQNPDEGTFEENLGKTA 90
Db 19 SFLGIPFAEPVGNLRFKAPQYKEPWSVDLDA---TKYPSCLDQDDDFGFSLSDL-KVA 74
Qy 91 LDLVMQSKVQAVLPQSEDCLTINVPPGPKAGANLPVLMIFGGGFEIGSPFIFFPAQ 150
Db 75 LKMLSLGWNKLVGLKJSEDCLYLVYTPKNTKPSKLPVWVIHGGGFMFGSGHSLPUL 134
Qy 151 MVTKSLVMGKHIIHVAVNYRVASWGLAGDDIIKAEGSGNAGLKQRLGMQWADNIAGFG 210
Db 135 YDGESLAREGNIVVSGINYLGLPLGFLSTGDDKLPGSGNYGLLDQRLALKWVQDNIAFG 194
Qy 211 GDPKSVTIFGESAGMSVLCILINWGDGNTYKGPLFRAGIMQSGA-----MVPSPDPVDG 265
Db 195 GDPNSVTIFGESAGAAVSLLLLISNGGDNPPSPKGLFHRAISQSGALSPPWAIQSESNAR 254
Qy 266 TYGNEIYDLFVSSAGCG---SNASDKLACLRASSDTLADATNNTPGFLAYS-----SLRSL 318
Db 255 GRAKELARLL---GCNETSSLELDCLRKSABELLEATRS---FLLFEVVPPLFLFLA 307
Qy 319 YLPRPDGKN-----ITDDMYKLVRDGYASVPVIIGQNDRGTTIFGLSSLVNTNAQAR- 372
Db 308 FGPVVDGDDAPEAFIPEDPELIKEGKFAVPYILGYTKDEGGYFAAMLLNASSKGEDEL 367
Qy 373 -----AYFKQSFHASDA-----EITLMAAYPQDITGSPFDTGVNALTTP 414
Db 368 KKETNPVWLELLKLYLFYASEALNIDKMDLADKVLKYPGDVDDFS-----YES 418
Qy 415 QPKISAVLGDLAF-----IHARR-----YFLNHFQGTGKYSLSKOLSLGSLPIM 458
Db 419 RKPQLQDMLTDLPLKCPTRVAADLHAKHGGSPVYAYFDHPASFGIGQFLAKRVDP-BFG 477
Qy 459 GTFHANDI--VWQDYL-----GSGSVIYNNAFIATFDLPDNP---TAGLLVNW 502
Db 478 GAVGDEIRFFVFGNPLKQELYKATEBEEKSSSKTMWYANFAKTGNPNNGTSNGLVW 537
Qy 503 PKYTSSQSGNNLMINAL 521
Db 538 PKYTSEQKSLILLATTI 556

RESULT 3
US-08-348-920-1
; Sequence 1, Application US/08348920
; Patent No. 5695750
; GENERAL INFORMATION:
; APPLICANT: Doctor, Bhupendra P.
; APPLICANT: Maxwell, Donald
; APPLICANT: Saxena, Ashima
; APPLICANT: Radic, Zoran
; APPLICANT: Taylor, Palmer
; TITLE OF INVENTION: Compositions for Use to Deactivate
; TITLE OF INVENTION: Organophosphates
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John F. Moran
; STREET: Off. of Command Judge Adv., HQ USAMRDC, Fort
; CITY: Detrick
; STATE: MD
; COUNTRY: US
; ZIP: 21702-5012
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
```

```
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/348,920
/ FILING DATE: 25-NOV-1994
/ CLASSIFICATION: 424
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Hendricks, Glena
/ REGISTRATION NUMBER: 32,535
/ REFERENCE/DOCKET NUMBER: doc348,920
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (301) 619-7807
/ TELEFAX: 301-619-7714
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 575 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: unknown
/ MOLECULE TYPE: protein
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ FRAGMENT TYPE: internal
/
US-08-348-920-1

Query Match 17.2%; Score 491.5; DB 1; Length 575;
Best Local Similarity 30.4%; Pred. No. 2.2e-39;
Matches 164; Conservative 75; Mismatches 195; Indels 105; Gaps 24;

QY 22 GDTITGLNAINFAFGIPFAEPVGNLRFKDP---VPYSGSLNGKQFTSYGSCMQNP 78
Db 17 GTRVPVLSHSI8-AFUGIPFAEPVGNMFRPRPEPKKPSWGVWNA5TY-----PNNCQQYV 71
QY 79 EGTFEENLGTALDLVMQSKVFQAVLPQSEDCLITINV---VRPPGTKAGANLPVMLWIF 134
Db 72 DEQFGFGSG-----SEMNPNREMSDCLYLNIVWPSRPKST-----TVMVWLY 116
QY 135 GGGFEIGSTIIPPAOMVTKSVLMGXH-----IIHVAVNYRVASWGFIA--GDDTKAEG 186
Db 117 GGGFYSGSSTL-----DVYNGKYLAYTEWVLVSLSYRVGAFGLALHGSQ---EA 164
QY 187 SGNAGLKDORLGOWADNIAGGDPKSVTIIFGESAGSMSVLCILWINDGNTYKGPL 246
Db 165 PGNVGLDQRMALQWHDNIQFFGGDPKVTIIFGESAGGASVGMHIL-----SPGSRDL 218
QY 361 -----SSLNVTNAQARAYFKQSFHASDAEIDTLMAAYPQDITQSSPDTGVNLAL 412
Db 334 YGAPGSKDSEKISREDPMWGVKLSVPRIANDLGLDAVTLQY-----TDMWDDNNGIKNR- 388
QY 413 TPQFKRISAVLGLDAFIHARRYFLN---HFQGGTKYSFLSKQLSGL---PIMGTFHANI 466

RESULT 4
US-08-348-920-2
/ Sequence 2, Application US/08348920
/ Patent No. 5695750
/ GENERAL INFORMATION:
/ APPLICANT: Doctor, Bhupendra P.
/ APPLICANT: Maxwell, Donald
/ APPLICANT: Saxena, Ashima
/ APPLICANT: Radic, Zoran
```

```
/ APPLICANT: Taylor, Palmer
/ TITLE OF INVENTION: Compositions for Use to Deactivate
/ NUMBER OF SEQUENCES: 2
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: John F. Moran
/ STREET: Off. of Command Judge Adv., HQ USAMRDC, Fort
/ CITY: Frederick
/ STATE: MD
/ COUNTRY: US
/ ZIP: 21702-5012
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ FILING DATE: 25-NOV-1994
/ CLASSIFICATION: 424
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Hendricks, Glena
/ REGISTRATION NUMBER: 32,535
/ REFERENCE/DOCKET NUMBER: doc348,920
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (301) 619-7807
/ TELEFAX: 301-619-7714
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 575 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: unknown
/ MOLECULE TYPE: protein
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ FRAGMENT TYPE: internal
/
US-08-348-920-2

Query Match 17.1%; Score 488.5; DB 1; Length 575;
Best Local Similarity 30.2%; Pred. No. 4.4e-39;
Matches 163; Conservative 76; Mismatches 195; Indels 105; Gaps 24;

QY 22 GDTITGLNAINFAFGIPFAEPVGNLRFKDP---VPYSGSLNGKQFTSYGSCMQNP 78
Db 17 GTRVPVLSHSI8-AFUGIPFAEPVGNMFRPRPEPKKPSWGVWNA5TY-----PNNCQQYV 71
QY 79 EGTFEENLGTALDLVMQSKVFQAVLPQSEDCLITINV---VRPPGTKAGANLPVMLWIF 134
Db 72 DEQFGFGSG-----SEMNPNREMSDCLYLNIVWPSRPKST-----TVMVWLY 116
QY 135 GGGFEIGSTIIPPAOMVTKSVLMGXH-----IIHVAVNYRVASWGFIA--GDDTKAEG 186
Db 117 GGGFYSGSSTL-----DVYNGKYLAYTEWVLVSLSYRVGAFGLALHGSQ---EA 164
QY 187 SGNAGLKDORLGOWADNIAGGDPKSVTIIFGESAGSMSVLCILWINDGNTYKGPL 246
Db 165 PGNVGLDQRMALQWHDNIQFFGGDPKVTIIFGESAGGASVGMHIL-----SPGSRDL 218
QY 247 FRAGIMQSGAMVPSDP---VDGTGNGEYIDLFFVSSAGCGSASDK--LACLRSSASTLLD 301
Db 219 FRRALIQSGS--PNCFWASVSVAEGRRAVELGRNLNCLNLSDEELIHLREKKPQELID 276
QY 302 ATNNTPGFLAYSSL-RLSYLPRDPKNIITDDMYKLVDRGKYASVPVILIGQNDDEGTIFGL 360
Db 277 VEWNV---LPFDSIFRFSFVPVIDGFEFFPTSLESMLNSGNFKKTKILLGVNKDEGSFFLL 333
QY 361 -----SSLNVTNAQARAYFKQSFHASDAEIDTLMAAYPQDITQSSPDTGVNLAL 412
Db 334 YGAPGSKDSEKISREDPMWGVKLSVPRIANDLGLDAVTLQY-----TDMWDDNNGIKNR- 388
QY 413 TPQFKRISAVLGLDAFIHARRYFLN---HFQGGTKYSFLSKQLSGL---PIMGTFHANI 466
```


REGISTRATION NUMBER: 38,972
REFERENCE/DOCKET NUMBER: 2391.00082
TELECOMMUNICATION INFORMATION:
TELEPHONE: (248) 539-5050
TELEFAX: (248) 539-5055
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 600 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-975-084-1

Query Match 15.8%; Score 452; DB 3; Length 600;
Best Local Similarity 29.1%; Pred. No. 1.8e-35;
Matches 156; Conservative 83; Mismatches 179; Indels 118; Gaps 26;
QY 35 AFLGIPPAEPVGNLRPKDP--VPYSGSLMGKQFTSYGPGCMQ-----QNPEGTFEEN 85
Db 62 AFLGIPPAEPVGNLRPKDP--VPYSGSLMGKQFTSYGPGCMQ-----QNPEGTFEEN 118
QY 86 LGKTALDLMQSKVFOAVLPQSEDCLTINVRPFGTKAGANLPVLMWIFGGGFEIGSPTI 145
Db 119 PNREL-----SEDCLYLN-VMTVPYRPTSPTPVLVWIIYGGGFYSGASSL 161
QY 146 PPPQAMVTKVLMGKHIIH-----VAVNRYVASWGLAGDDIIKAGSGNAGLQDRLGM 199
Db 162 -----DYYDGRFLVQAERTVLVSMNRYVGAFGLALPGSR-EAPGNVGLLDQRLAL 211
QY 200 QWADNLAGFGGDPKVTIFGSGAGSMVLCILWINDGNTYKGLPLFRAGIMOSGAMVP 259
Db 212 QWQENVAAGGDPSTVTLFGSAGAAVSGLHLL-----SPPERGLEHRAVLQSGA--P 263
QY 260 SDPVDGTYG-----NEIYDLFVSSAGC-----GSADKLACLRSSASDITLDAFN 305
Db 264 NGPW-ATVGMGEARRATQLAHL-----VGCPPGGTGGNDTELVACLRTPAQVLV---NH 315
QY 306 TPGFLAYSSL-RLSYLPRPDGKNTDDMYKLVROGKYASVPIIGDQNDGTFI----- 358
Db 316 EHWVLPQSSVFRFSFVPPVVDGFLSDTPEALINAGDFHGLQVLGVGVKDEGSYLVYGA 375
QY 359 GLSSLN--VTTNAQARAYFKQSFIHASDAEIDTLMAAY-----POD-----ITQGGPFDTVG 408
Db 376 GFSKDNESLISRAEFLAGRVGVQVSDLAABAVVLHYTDMLHPEDPARLEALSDVVG 435
QY 409 LNALTPQPKRISAVLGLAFIAR--RYFLNHFGQ-----GTKYSFLSKQLSGLPI- 457
Db 436 HNVVCF-----VAQLAGRLAAQGARVYAVVFEHRASTLSWPLMGMVPHGYEIEFIFGIPLD 491
QY 458 -MGTFHNDIWDQYLLGSGSVIYNNAFIATATDLPNTA--GLLVNWPKYTSSSQ 510
Db 492 PSRNYTAEKIFAQRLM-----RYWANFARTGDPNEPRDPKAPQWPPYTAGAQ 539

RESULT 9

5215909-11
; Patent No. 5215909
; APPLICANT: SOREX, HERMONA
; TITLE OF INVENTION: HUMAN CHOLINESTERASE GENES
; NUMBER OF SEQUENCES: 13
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/572,911
; FILING DATE: 15-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 87,724
; FILING DATE: 21-AUG-1987
; APPLICATION NUMBER: 875,737
; FILING DATE: 18-JUN-1986
; SEQ ID NO: 11
; LENGTH: 602

5215909-11

Query Match 15.8%; Score 452; DB 6; Length 602;
Best Local Similarity 29.5%; Pred. No. 1.9e-35;
Matches 158; Conservative 82; Mismatches 150; Indels 106; Gaps 25;
QY 22 GDTITGLNATINEAFLGIPFABPVGNLRPKDPVYPYSGSLMGKQFTSYGSGCMQNEPST 81
Db 50 GGTVT-----AFLGIPYAAPPLGRLEKPKQSLTKWSDIWNATKYANSCC-QNIDQS 100
QY 82 FEENLGTALDLMQSKVFOAVLPQSEDCLTINVRP-PGTKAGANLPVLMWIFGGGFEI 140
Db 101 PFGFHG-----SEMNPNPTDLSDCLYLNVWIPAPKPK---NATVLIWIIYGGGFT 148
QY 141 GSPTIFPPAQMVTKVLMGK-----HIIHVAVNRYVASWGLAGDDIIKAGSGNAGLKD 194
Db 149 GTSSL-----HVVYDGKFLARVERVIVVSMNRYVGALGFLAWPG-NPEAPGNMGLFD 198
QY 195 QRLGMQWADNLAGFGGDPKVTIFGSGAGSMVLCILWINDGNTYKGLPLFRAGIMOS 254
Db 199 QQLALQWQKNIAAFAGGNPKSVTLFGSAGAAVSGLHLL-SPGSHS-----LFTAILQS 252
QY 255 GAMVPSDPVDGTY--GNEIYDLFVSSAGC--SASDKLACLRSSASDITLDAFNTPGFL 310
Db 253 GSNAPWAVTSLYEARNTLNL-AKLTGCSRENTEIILKLRNKDPOEIL---LINEAFV 308
QY 311 AYSS-LRLSYLPRPDGKNTDDMYKLVROGKYASVPIIGDQNDGTFI-----GLSSL 363
Db 309 PYGTPLSVNFGPTVDGFLDIMPDLILLELGGQFKTKQILVGVNKDEGTAFLVYGAPGFSK 368
QY 364 N--VTTNAQARAYFKQSFIHASDAEIDTLMAAY-----PODITQSGPFDTVGLNALT 413
Db 369 NNSIITRKEFOGLKIFPPGVSEFGKESILFHYTDWVDQRPENYREALGDVWGDYDNFIC 428
QY 414 P-QFKRISAVLGLAFIARRYFLNHFGQTKYSFLSKQLSGLPIMCTPHANDIWMODY 471
Db 429 PALBETKFKSEWGNAPF---YIFEH-----RSSKLPPWFEGWGMGYEI---EF 472
QY 472 LLG-----SGSVIYNNAFIATATDLPN-TAGLLVNWPKYTSSTSSQ 510
Db 473 VFGLPLERRDNYTKAEILSHSIV--KRWANFAKYGNPNETQNNSTSWPVPFKSTEQ 526

RESULT 10

US-07-732-962A-2
; Sequence 2, Application US/07732962A
; Patent No. 5248604
; GENERAL INFORMATION:
; APPLICANT: Fischer, Meir
; TITLE OF INVENTION: EXPRESSION OF ENZYMATICALLY ACTIVE
; TITLE OF INVENTION: RECOMBINANT HUMAN ACETYLCHOLINESTERASE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White, Esq.
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/732,962A
; FILING DATE: 19910722
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 39304/JPW/LSW
; TELECOMMUNICATION INFORMATION:

```
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 614 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-732-962A-2

Query Match      15.8%; Score 452; DB 1; Length 614;
Best Local Similarity 29.1%; Pred. No. 1.9e-35;
Matches 156; Conservative 83; Mismatches 179; Indels 118; Gaps 26;

QY 35 AFLGIPFAEPVGNLRFKDP---VPYSGSLNGKQKFTSYGSPCMQ-----QNPEGTFEEN 85
Db 62 AFLGIPFAEPVGNLRFKDP---VPYSGSLNGKQKFTSYGSPCMQ-----QNPEGTFEEN 118
QY 86 LGKTALDLVMQSKVFOAVLPQSEDCLTINNVPRPGTKAGANLPVLMWIFGGGFEIGSPTI 145
Db 119 PNREL-----SEDCLYLN-VWTPYPRPTSPTPLVLMWYGGGFEIGSGASSL 161
QY 146 FPPAQMTKSVLMGXHIIH-----VAVNYRVASWGFLAGDDIKABGSGNAGLKDORLGM 199
Db 162 -----DVIDGRFLVQAERTVLVSMNYRVGAFGLALPGSR-EAPGNVGLLDORLAL 211
QY 200 QWVADNIAGFGDPSKVTIFGESAGSMSVLCILWNGDNTYKGLPLFRAGIMQSGMVP 259
Db 212 QWQVENVAFAFGDPTSVTLFESAGASVGMHLL-----SPPSRGLFHRAVLQSGA--P 263
QY 260 SDPVDGTYG-----NEIYDLFVSSAGC-----GSASDKLACLRSSASDTLLDATNN 305
Db 264 NGPW-ATVGMGARRRATQLAHL-----VCGPPGCTGNDTELVAQLTRPAQVLV---NH 315
QY 306 TPGFLAYSSL-RLSYLPRPDGKNIITDDMYKLVRDGYASVPVLIIGQNDDEGTIF----- 358
Db 316 EWHVLPOESVFRFSFVVDGDFLSDTPALINAGDFHGLQVLGVVKGDEGSFVLVYGAP 375
QY 359 GLSSLN--VTTNAQARAYFKQSFIIHASDAEIDTLMAAY-----PQD---ITQSPDPTGV 408
Db 376 GFSKONESLISRAEFLAGVRGVGPQVSDLAEEAVLHYTDWLHPEDPARLREALSDVVDG 435
QY 409 LNALTPQFKRISAVLDGLAFIAR--RYFLNHFGQ-----GTKYSFLSKOLSGLPI- 457
Db 436 HNVVCP-----VAQLAGRLAAQGARVYVYVFEHRASTLSWPLWGVPHGVEIEFIFGIPLD 491
QY 458 -MGTTHANDIVQDYLILGSGSVIYNNAFIAFATDLDNPTA--GLLVNWPKYTSSSQ 510
Db 492 PSRNYTAEEKIFAQRLM-----RYWANFARTGDPNEPRDPKAPQWPPYTAGAQ 539

RESULT 11
US-08-370-156-2
; Sequence 2, Application US/08370156
; Patent No. 5932780
; GENERAL INFORMATION:
; APPLICANT: Soreq, Hermona
; APPLICANT: Zakut, Haim
; APPLICANT: Shani, Moshe
; TITLE OF INVENTION: TRANSGENIC ANIMAL ASSAY SYSTEM FOR
; TITLE OF INVENTION: ANTICHOLINESTERASE SUBSTANCES
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Reising, Rethington, Barnard & Perry
; STREET: P.O. Box 4390
; CITY: Troy
; STATE: Michigan
; COUNTRY: US
; ZIP: 48099
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/370,156
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,955
; REFERENCE/DOCKET NUMBER: P-307 (Mulford)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 689-3500
; TELEFAX: (810) 689-4071
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 614 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-370-156-2

Query Match      15.8%; Score 452; DB 2; Length 614;
Best Local Similarity 29.1%; Pred. No. 1.9e-35;
Matches 156; Conservative 83; Mismatches 179; Indels 118; Gaps 26;

QY 35 AFLGIPFAEPVGNLRFKDP---VPYSGSLNGKQKFTSYGSPCMQ-----QNPEGTFEEN 85
Db 62 AFLGIPFAEPVGNLRFKDP---VPYSGSLNGKQKFTSYGSPCMQ-----QNPEGTFEEN 118
QY 86 LGKTALDLVMQSKVFOAVLPQSEDCLTINNVPRPGTKAGANLPVLMWIFGGGFEIGSPTI 145
Db 119 PNREL-----SEDCLYLN-VWTPYPRPTSPTPLVLMWYGGGFEIGSGASSL 161
QY 146 FPPAQMTKSVLMGXHIIH-----VAVNYRVASWGFLAGDDIKABGSGNAGLKDORLGM 199
Db 162 -----DVIDGRFLVQAERTVLVSMNYRVGAFGLALPGSR-EAPGNVGLLDORLAL 211
QY 200 QWVADNIAGFGDPSKVTIFGESAGSMSVLCILWNGDNTYKGLPLFRAGIMQSGMVP 259
Db 212 QWQVENVAFAFGDPTSVTLFESAGASVGMHLL-----SPPSRGLFHRAVLQSGA--P 263
QY 260 SDPVDGTYG-----NEIYDLFVSSAGC-----GSASDKLACLRSSASDTLLDATNN 305
Db 264 NGPW-ATVGMGARRRATQLAHL-----VCGPPGCTGNDTELVAQLTRPAQVLV---NH 315
QY 306 TPGFLAYSSL-RLSYLPRPDGKNIITDDMYKLVRDGYASVPVLIIGQNDDEGTIF----- 358
Db 316 EWHVLPOESVFRFSFVVDGDFLSDTPALINAGDFHGLQVLGVVKGDEGSFVLVYGAP 375
QY 359 GLSSLN--VTTNAQARAYFKQSFIIHASDAEIDTLMAAY-----PQD---ITQSPDPTGV 408
Db 376 GFSKONESLISRAEFLAGVRGVGPQVSDLAEEAVLHYTDWLHPEDPARLREALSDVVDG 435
QY 409 LNALTPQFKRISAVLDGLAFIAR--RYFLNHFGQ-----GTKYSFLSKOLSGLPI- 457
Db 436 HNVVCP-----VAQLAGRLAAQGARVYVYVFEHRASTLSWPLWGVPHGVEIEFIFGIPLD 491
QY 458 -MGTTHANDIVQDYLILGSGSVIYNNAFIAFATDLDNPTA--GLLVNWPKYTSSSQ 510
Db 492 PSRNYTAEEKIFAQRLM-----RYWANFARTGDPNEPRDPKAPQWPPYTAGAQ 539

RESULT 12
US-08-446-100-19
; Sequence 19, Application US/08446100
; Patent No. 6001625
; GENERAL INFORMATION:
; APPLICANT: Broomfield, Clarence A
; APPLICANT: Millard, Charles B
; APPLICANT: Lockridge, Oksana
; TITLE OF INVENTION: Site-Directed Mutagenesis of Esterases
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
```

ADDRESSEE: Hendricks and Assoc.
STREET: 9669 A Main Street, P.O. Box 2509
CITY: Fairfax
STATE: VA
COUNTRY: US
ZIP: 22031
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446.100
FILING DATE: 19-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hendricks, Glenna
REGISTRATION NUMBER: 32,535
REFERENCE/DOCKET NUMBER: broomfield
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 425-4250
TELEFAX: (703) 425-2767
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 614 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: YES
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE: human esterases
US-08-446-100-19

Query Match 15.8%; Score 452; DB 3; Length 614;
Best Local Similarity 29.1%; Pred. No. 1.9e-35;
Matches 156; Conservative 83; Mismatches 179; Indels 118; Gaps 26;
QY 35 AFLGIPFAEPVGNLRFKDP---VPYSGSLNGQKFTSYGSPCMQ-----QNPGTFTEEN 85
DB 62 AFLGIPFAEPVGNLRFKDP---VPYSGSLNGQKFTSYGSPCMQ-----QNPGTFTEEN 118
QY 86 LGKLTALDLVMSKVFQAVLPQSEDCLTINVRPPTKAGANLPVWLITFGGFEIGSPTI 145
DB 119 PNREL-----SEDCLYLN-VMTYPRPTSPVPLVWYGGFGYSGASSL 161
QY 146 FPPAQMTKSVLMGKHIIH-----VAVNYRVASWGLAGDDIKAEGSGNAGLKDQRLGM 199
DB 162 -----DVYDGRFLVQAERTVLVSMNRYVAGFGLALPGSR-EAPGNVGLLDQRLAL 211
QY 200 QWADNIAGFGDPSKVTIFGESAGSMVLCILWINDGNTYKGPPLFRAGIMOSGAMVP 259
DB 212 QWQENVAAGFGDPTSVTLFGESAGASVGMHLL-----SPPSRGLFHRAVLQSGA--P 263
QY 260 SDPVDGTG-----NEIYDLFVSSAGC-----GSASDKLACLRASDITLDATNN 305
DB 264 NGPW-ATVGMGEARRATQLAHL-----VGCPPGGTGGNDTELVACLRTFPAQLV--NH 315
QY 306 TPGFLAYSSL-RLSYLPRPDGKNITDDMYKLVROGKYASVPVIGDQNDGTIF----- 358
DB 316 EHWLVPQESVFRFSFVPVVDGDFLSDTPEALINAGDFHGLQVLGVVVKDEGSYFLVYAP 375
QY 359 GLSSIN-VTTNAQARVFKOSFIHASDAEIDTLMAAY-----POD-----ITQSPFDGV 408
DB 376 GFSKDNESLIRABFLAGRVGVVPOVSDLAEEAVVLHYDMLHDPEDPARLREALSDVVGD 435
QY 409 LNALTPQPKRSAYLGLDLAFIAR--RYFLNHFGQ-----GTKYFSLKQSLGLPI- 457
DB 436 HNVVCP-----VAQLAGLAAQAGRYAVVFEHRASTLSMPLWMGVPHGYEIEFIFIGILD 491
QY 458 -MGTHANDIVQDYLLGSGSVIYNNAFIATDLDPNTA--GLLVNWPKYTSSSQ 510

Db 492 PSRNYTAEKIFAQRLM-----RYWANPARTGDPNEPRDKAPQWPPYTAGAQ 539
RESULT 13
US-08-446-100-21
Sequence 21, Application US/08446100
Patent No. 6001625
GENERAL INFORMATION:
APPLICANT: Broomfield, Clarence A
APPLICANT: Millard, Charles B
APPLICANT: Lockridge, Oksana
TITLE OF INVENTION: Site-Directed Mutagenesis of Esterases
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hendricks and Assoc.
STREET: 9669 A Main Street, P.O. Box 2509
CITY: Fairfax
STATE: VA
COUNTRY: US
ZIP: 22031
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446.100
FILING DATE: 19-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hendricks, Glenna
REGISTRATION NUMBER: 32,535
REFERENCE/DOCKET NUMBER: broomfield
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 425-4250
TELEFAX: (703) 425-2767
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 614 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: YES
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE: human esterases
US-08-446-100-21
Query Match 15.8%; Score 452; DB 3; Length 614;
Best Local Similarity 29.1%; Pred. No. 1.9e-35;
Matches 156; Conservative 83; Mismatches 179; Indels 118; Gaps 26;
QY 35 AFLGIPFAEPVGNLRFKDP---VPYSGSLNGQKFTSYGSPCMQ-----QNPGTFTEEN 85
DB 62 AFLGIPFAEPVGNLRFKDP---VPYSGSLNGQKFTSYGSPCMQ-----QNPGTFTEEN 118
QY 86 LGKLTALDLVMSKVFQAVLPQSEDCLTINVRPPTKAGANLPVWLITFGGFEIGSPTI 145
DB 119 PNREL-----SEDCLYLN-VMTYPRPTSPVPLVWYGGFGYSGASSL 161
QY 146 FPPAQMTKSVLMGKHIIH-----VAVNYRVASWGLAGDDIKAEGSGNAGLKDQRLGM 199
DB 162 -----DVYDGRFLVQAERTVLVSMNRYVAGFGLALPGSR-EAPGNVGLLDQRLAL 211
QY 200 QWADNIAGFGDPSKVTIFGESAGSMVLCILWINDGNTYKGPPLFRAGIMOSGAMVP 259
DB 212 QWQENVAAGFGDPTSVTLFGESAGASVGMHLL-----SPPSRGLFHRAVLQSGA--P 263
QY 260 SDPVDGTG-----NEIYDLFVSSAGC-----GSASDKLACLRASDITLDATNN 305

Db 264 NGPW-ATVGMGARRRATQLAHL-----VGPDPGGTGGNDTELVAQLTRTPAQVLLV---NH 315
QY 306 TPGFLAYSSL-RLSYLPRDGNKNIITDDMYKLVDRDGYASVVPVLIQDNDGTFIF----- 358
Db 316 EHWLVLPQESVHRFSFVVPVVDGFLSTPEALINAGDFHGLQVLGVVKGDEGSFLVYGAP 375
QY 359 GLSSLN--VTTNAQARAYFKQSFIIHASDAEIDTLMAAY-----PQD---ITQSPDPDTGV 408
Db 376 GFSDKNESSLISRAEFLAGVRGVGPQVSDLAEEAVVLHYTDLHPEDPARLREALSDVVG 435
QY 409 LNALTPQFKRISAVLGDIAFIHAR--RYFLNHPQG-----GTKYSLFKSLQSLGLPI- 457
Db 436 HNVVCP-----VAQLAGRLAAQAGRVYVYFHEHRASTLSWPLMWGVPHGYEIEFIFGIPLD 491
QY 458 -MGTTHANDIVWQDVLGSGSVIYNNAFIATFATDLDNTA--GLLVNWPKYTSSSQ 510
Db 492 PSRNTAEKIFAQRLM-----RYWANFARTGDPNEPRDPKAPQWPPYTAGAQ 539

RESULT 14

US-08-814-095-2
; Sequence 2, Application US/08814095
; Patent No. 6025183
; GENERAL INFORMATION:
; APPLICANT: Soreq, Hermona
; APPLICANT: Zakut, Haim
; APPLICANT: Shani, Moshe
; TITLE OF INVENTION: TRANSGENIC ANIMAL ASSAY SYSTEM FOR
; TITLE OF INVENTION: ANTI-CHOLINESTERASE SUBSTANCES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KOHN & ASSOCIATES
; STREET: 30500 No. 6025183thwestern Highway, Suite 410
; CITY: Farmington Hills
; STATE: Michigan
; COUNTRY: U.S.
; ZIP: 48334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/814,095
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Montgomery, Ilene N.
; REGISTRATION NUMBER: 38,972
; REFERENCE/DOCKET NUMBER: 2391.00066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (248) 539-5050
; TELEFAX: (248) 539-5055

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 614 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-814-095-2

Query Match 15.8%; Score 452; DB 3; Length 614;
Best Local Similarity 29.1%; Pred. No. 1.9e-35;
Matches 156; Conservative 83; Mismatches 179; Indels 118; Gaps 26;

QY 35 AFLGIFPAPPVGNLRFKDP---VPYSGSLNGQKETSYPSCMQ-----QNPEGTFEEN 85
Db 62 AFLGIFAPGPPGRRFLPEPKQPSGVVDATTFQSV---CQYVDTLYPGEFEGEMWN 118
QY 86 LGKTALDLYMQSKVFOAVLQSEDDCLTINVVRPPTGKAGANLPVLMWIFGGGFEIGSPTI 145

Db 119 PNREL-----SEDCLYLN-VWTPYRPTSPTPVLVWLYGGGFYSGASSL 161
QY 146 FPPAQMTKSVLMKGHIH-----VAVNYRVASGFLAGDDIKAGSGNAGLKDORLGM 199
Db 162 -----DVTYDGRFLVQAERTVLVSMNRYRGAFGFALPGSR-EAPGNVGLLDORLAL 211
QY 200 QWVADNIAFGGDPKSVTIIFGESAGSMSVLCHLIWNDDGNTYKGLPFLFRAGIMQSGAMVP 259
Db 212 QWVQENVAFAFGDPTSVTLFSGSAGASVGMHL-----SPSRGLFHRAVLQSGA--P 263
QY 260 SDPVDTGY-----NEIYDLFVSSAGC-----GSASDKLACLRASSDITLLDATNN 305
Db 264 NGPW-ATVGMGARRRATQLAHL-----VGPDPGGTGGNDTELVAQLTRTPAQVLLV---NH 315
QY 306 TPGFLAYSSL-RLSYLPRDGNKNIITDDMYKLVDRDGYASVVPVLIQDNDGTFIF----- 358
Db 316 EHWLVLPQESVHRFSFVVPVVDGFLSTPEALINAGDFHGLQVLGVVKGDEGSFLVYGAP 375
QY 359 GLSSLN--VTTNAQARAYFKQSFIIHASDAEIDTLMAAY-----PQD---ITQSPDPDTGV 408
Db 376 GFSDKNESSLISRAEFLAGVRGVGPQVSDLAEEAVVLHYTDLHPEDPARLREALSDVVG 435
QY 409 LNALTPQFKRISAVLGDIAFIHAR--RYFLNHPQG-----GTKYSLFKSLQSLGLPI- 457
Db 436 HNVVCP-----VAQLAGRLAAQAGRVYVYFHEHRASTLSWPLMWGVPHGYEIEFIFGIPLD 491
QY 458 -MGTTHANDIVWQDVLGSGSVIYNNAFIATFATDLDNTA--GLLVNWPKYTSSSQ 510
Db 492 PSRNTAEKIFAQRLM-----RYWANFARTGDPNEPRDPKAPQWPPYTAGAQ 539

RESULT 15

PCT-US92-06106-2
; Sequence 2, Application PC/TUS9206106
; GENERAL INFORMATION:
; APPLICANT: Fischer, Meir
; TITLE OF INVENTION: ENZYMATICALLY ACTIVE RECOMBINANT HUMAN
; TITLE OF INVENTION: ACETYLCHOLINESTERASE AND USES THEREOF
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White, Esq.
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06106
; FILING DATE: 19920722
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 39304-B-PCT/JPW/EAB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 614 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-06106-2

Query Match 15.8%; Score 452; DB 5; Length 614;

Search completed: July 29, 2004, 10:51:30
Job time : 20 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 10:42:58 ; Search time 55 Seconds
(without alignments)
2810.063 Million cell updates/sec

Title: US-09-943-857-4
Perfect score: 2864
Sequence: 1 SMNSRGPAAGRLGVSPTAKIA.....DNFRTAGYDALMTNPSFFV 547

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04: *
1: Geneseqp1980s: *
2: Geneseqp1990s: *
3: Geneseqp2000s: *
4: Geneseqp2001s: *
5: Geneseqp2002s: *
6: Geneseqp2003as: *
7: Geneseqp2003bs: *
8: Geneseqp2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2864	100.0	547	6	ABU09071 Candida r
2	2588	90.4	547	6	ABU09074 Candida r
3	2495	87.1	547	6	ABU09073 Candida r
4	2459	85.9	534	4	AB47622 Synthetic
5	2434	85.0	547	6	ABU09072 Candida r
6	2396	83.7	547	6	ABU09070 Candida r
7	1107	38.7	563	2	AAR47577 Lipase of
8	1097	38.3	563	2	AAR10330 Gene prod
9	1078	37.6	557	2	AAR06370 Protein w
10	674.5	23.6	584	5	ABB07676 Carboxyle
11	685	23.2	575	5	ABB79538 Carboxyle
12	609	21.3	571	4	ABU53224 Human met
13	491.5	17.2	575	2	AAW39078 Torpedo c
14	488.5	17.1	575	2	AAW39079 Torpedo c
15	484	16.9	576	7	ABR84598 T califon
16	482	16.8	574	5	AAO18905 Rat butyr
17	477	16.7	574	5	AAO18903 Horse but
18	469	16.4	614	7	ADE61695 Rat Prote
19	469	16.4	614	7	ADE61687 Rat Prote
20	469	16.4	614	7	ADE61683 Rat Prote
21	469	16.4	614	7	ADE61691 Rat Prote
22	467	16.3	614	6	ABP59222 Human dru
23	462	16.1	826	2	AAV30100 Amino aci
24	462	16.1	826	2	AAV30101 An acetyl
25	459	16.0	574	5	AAO18904 Cat butyr

26	455	15.9	574	7	ABW00725 Human but
27	454	15.9	614	3	AAV49495 Human ace
28	454	15.9	614	3	AAV49494 Human ace
29	452	15.8	583	3	AG80773 AchE prot
30	452	15.8	600	2	AAW48797 Human ace
31	452	15.8	602	2	AAR37442 Full-leng
32	452	15.8	614	2	AAR80726 Human ace
33	452	15.8	614	3	AAV49489 Human wil
34	452	15.8	614	3	AAV49491 Human ace
35	452	15.8	614	5	AAU11231 Human ace
36	452	15.8	614	5	AAU11232 Human ace
37	452	15.8	614	5	AAU11234 Human ace
38	452	15.8	614	6	ABB99726 Amino aci
39	452	15.8	614	7	ADE61697 Human Pro
40	452	15.8	614	7	ADE61689 Human Pro
41	452	15.8	614	7	ADE61693 Human Pro
42	452	15.8	614	7	ADE61685 Human Pro
43	452	15.8	620	5	AAU11235 Human ace
44	451	15.7	574	5	AAE25235 Human but
45	451	15.7	574	5	AAO18898 Human but

ALIGNMENTS

RESULT 1
ABU09071
ID ABU09071 standard; protein; 547 AA.
XX

AC ABU09071;
XX
15-JUL-2003 (first entry)
XX

DE Candida rugosa lipase 3 polypeptide.
XX
KW Lipase 3; enzyme; mutant; muten.
XX

OS Candida rugosa.
XX
Key Location/Qualifiers
FT Misc-difference 1..13

FT /note= "Absent in wild-type Candida rugosa lipase"
FT Misc-difference 14
FT /note= "Wild-type Ala substituted by Val"

FT Misc-difference 161
FT /note= "Wild-type Pro substituted by His"

FT Misc-difference 408
FT /note= "Wild-type Ile substituted by Val"

FT Misc-difference 409
FT /note= "Wild-type Phe substituted by Leu"

FT Misc-difference 412
FT /note= "Wild-type Ile substituted by Leu"

EP1288294-A2.
XX

05-MAR-2003.
XX

26-APR-2002; 2002EP-00009616.
XX

31-AUG-2001; 2001US-00943857.
XX

(SINI-) ACAD SINICA.
XX

Tang S, Lee G, Shaw J;
XX

WPI; 2003-395476/38.
XX

N-PSDB; ABX95906.
XX

Isolated mutant nucleic acid encoding Candida rugosa lipase, useful for the preparation of Candida rugosa lipase for biocatalytic applications.
PT
Claim 35; Page 5-7; 33pp; English.
PS
XX

CC The invention relates to an isolated mutant nucleic acid encoding a
 CC Candida rugosa lipase polypeptide. The DNA has a sequence having at least
 CC 80% identity to a wild-type DNA encoding Candida rugosa lipase. The
 CC invention also relates to a microorganism comprising the DNA, where the
 CC microorganism is a bacterium or yeast, preparing a mutant DNA encoding a
 CC C.rugosa lipase and a chimeric C.rugosa lipase comprising a substrate
 CC interacting domain of a first C.rugosa lipase and a non-substrate
 CC interacting domain of a second C.rugosa lipase. The method is useful for
 CC preparing a mutant DNA encoding a Candida rugosa lipase. The DNA is
 CC useful in the large scale manufacture of Candida rugosa lipase which is
 CC useful for biocatalytic applications. This sequence represents Candida
 CC rugosa lipase 3
 XX
 SQ Sequence 547 AA;

Query Match 100.0%; Score 2864; DB 6; Length 547;
 Best Local Similarity 100.0%; Pred. No. 4,8e-245;
 Matches 547; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 SMNSRGAGRLGVSPTAKLANGDITITGLNAINAEFLGIPPAEPVGNLRKDPVPYSGS 60
 Db 1 SMNSRGAGRLGVSPTAKLANGDITITGLNAINAEFLGIPPAEPVGNLRKDPVPYSGS 60
 Qy 61 LMGKFTSYGSCMQNPEGTFEENLGTALDLVMQSKVFOAVLPQSEDCLTINVRPPG 120
 Db 61 LMGKFTSYGSCMQNPEGTFEENLGTALDLVMQSKVFOAVLPQSEDCLTINVRPPG 120
 Qy 121 TKAGANLPVMLWIFGGFEGISPTIFPPAQWTKSVLMGKHIHVAVNYRVASGFLAGD 180
 Db 121 TKAGANLPVMLWIFGGFEGISPTIFPPAQWTKSVLMGKHIHVAVNYRVASGFLAGD 180
 Qy 181 DIKAGSGNAGLKDQRLQMWADVNIAGFGDPSKVTIFGSGMSVLCHLIWNGDNT 240
 Db 181 DIKAGSGNAGLKDQRLQMWADVNIAGFGDPSKVTIFGSGMSVLCHLIWNGDNT 240
 Qy 241 YKGKPLFRAGIMQSGAMVPSDPVDTGYGNEIYDLFVSSAGCGSASDKLCLRSASSTLL 300
 Db 241 YKGKPLFRAGIMQSGAMVPSDPVDTGYGNEIYDLFVSSAGCGSASDKLCLRSASSTLL 300
 Qy 301 DATNTTGFAYSSRLSYLPRPGKNTDDMYKLVRDGYASVPVIIGDNDGTFGL 360
 Db 301 DATNTTGFAYSSRLSYLPRPGKNTDDMYKLVRDGYASVPVIIGDNDGTFGL 360
 Qy 361 SSLNVTNQAQARAYFKQSFHASDAEIDTLMAYPQDITQGSPTGTGINALTQPKRIS 420
 Db 361 SSLNVTNQAQARAYFKQSFHASDAEIDTLMAYPQDITQGSPTGTGINALTQPKRIS 420
 Qy 421 AVLGDLPFIHARRYFLAHFQGGTKYSFLSKQLSGLPIMGTTHANDIVWQDYLLGSGSVIY 480
 Db 421 AVLGDLPFIHARRYFLAHFQGGTKYSFLSKQLSGLPIMGTTHANDIVWQDYLLGSGSVIY 480
 Qy 481 NNAFIAFATDLPNTAGLLVNWPKYTSQSGNNLMINALGLYTGKDNFTAGYDALMT 540
 Db 481 NNAFIAFATDLPNTAGLLVNWPKYTSQSGNNLMINALGLYTGKDNFTAGYDALMT 540
 Qy 541 NPSSFFV 547
 Db 541 NPSSFFV 547

RESULT 2
 ABU09074
 ID ABU09074 standard; protein; 547 AA.
 XX
 AC ABU09074;
 XX
 XX
 DT 15-JUL-2003 (first entry)
 DE Candida rugosa lipase 8 polypeptide.
 XX
 XX Lipase 8; enzyme; mutant; mutein.
 XX
 OS Candida rugosa.

XX Key Location/Qualifiers
 FH Misc-difference 1. .13
 FT Misc-difference /note= "Absent in wild-type Candida rugosa lipase"
 FT Misc-difference 14
 FT Misc-difference /note= "Wild-type Ala substituted by Val"
 FT Misc-difference 197
 FT Misc-difference /note= "Wild-type Leu substituted by Met"
 FT Misc-difference 266
 FT Misc-difference /note= "Wild-type Ile substituted by Val"
 FT Misc-difference 278
 FT Misc-difference /note= "Wild-type Asn substituted by Asp"
 FT Misc-difference 333
 FT Misc-difference /note= "Wild-type Tyr substituted by Phe"
 FT Misc-difference 343
 FT Misc-difference /note= "Wild-type Asn substituted by Ser"
 FT Misc-difference 344
 FT Misc-difference /note= "Wild-type Ile substituted by Val"
 FT Misc-difference 370
 FT Misc-difference /note= "Wild-type Gln substituted by Glu"
 FT Misc-difference 373
 FT Misc-difference /note= "Wild-type Glu substituted by Gln"
 FT Misc-difference 376
 FT Misc-difference /note= "Wild-type Lys substituted by Thr"
 FT Misc-difference 387
 FT Misc-difference /note= "Wild-type Ile substituted by Leu"
 FT Misc-difference 396
 FT Misc-difference /note= "Wild-type Gly substituted by Gln"
 FT Misc-difference 408
 FT Misc-difference /note= "Wild-type Ile substituted by Val"
 FT Misc-difference 427
 FT Misc-difference /note= "Wild-type Gly substituted by Ala"
 FT Misc-difference 429
 FT Misc-difference /note= "Wild-type Thr substituted by Ile"
 FT Misc-difference 430
 FT Misc-difference /note= "Wild-type Leu substituted by His"
 FT Misc-difference 530
 FT Misc-difference /note= "Wild-type Phe substituted by Ser"
 XX
 EP1288294-A2.
 XX
 PD 05-MAR-2003.
 XX
 XX 26-APR-2002; 2002EP-00009616.
 XX
 XX 31-AUG-2001; 2001US-00943857.
 XX
 XX (SINI-) ACAD SINICA.
 XX
 XX Tang S, Lee G, Shaw J;
 XX
 XX WPI; 2003-395476/38.
 XX N-PSDB; ABX95909.
 XX
 XX Isolated mutant nucleic acid encoding Candida rugosa lipase, useful for
 XX the preparation of Candida rugosa lipase for biocatalytic applications.
 XX
 XX Claim 35; Page 10-11; 33pp; English.
 XX
 XX The invention relates to an isolated mutant nucleic acid encoding a
 XX Candida rugosa lipase polypeptide. The DNA has a sequence having at least
 XX 80% identity to a wild-type DNA encoding Candida rugosa lipase. The
 XX invention also relates to a microorganism comprising the DNA, where the
 XX microorganism is a bacterium or yeast, preparing a mutant DNA encoding a
 XX C.rugosa lipase and a chimeric C.rugosa lipase comprising a substrate
 XX interacting domain of a first C.rugosa lipase and a non-substrate
 XX interacting domain of a second C.rugosa lipase. The method is useful for
 XX preparing a mutant DNA encoding a Candida rugosa lipase. The DNA is
 XX useful in the large scale manufacture of Candida rugosa lipase which is
 XX useful for biocatalytic applications. This sequence represents Candida
 XX rugosa lipase 8
 XX
 XX Sequence 547 AA;

```

Query Match      90.4%; Score 2588; DB 6; Length 547;
Best Local Similarity 88.7%; Pred. No. 1.5e-220;
Matches 485; Conservative 30; Mismatches 32; Indels 0; Gaps 0;

QY 1 SMNSRGPAGRLGSPVTAKLANGDTTGLNAINAEFLGIPFAEPVGNLRFKDPVPYSGS 60
DB 1 SMNSRGPAGRLGSPVTATLANGDTTGLNAINAEFLGIPFAEPVGNLRFKDPVPYSGS 60

QY 61 LINGQKFTSYGSCMQNPGTEENLGKTDALDLMQSKVFOAVLPQSEDLTINVRPFG 120
DB 61 LINGQKFTSYGSCMQNPGTEENLGKTDALDLMQSKVFOAVLPQSEDLTINVRPFG 120

QY 121 TKAGANLPVLMWIFGGGFEIGSPTIPPPAQMYTKSVLMGKHIIHVAVNYRVASWGFLAGD 180
DB 121 TKAGANLPVLMWIFGGGFEIGSPTIPPPAQMYTKSVLMGKHIIHVAVNYRVASWGFLAGD 180

QY 181 DIKAEAGSNAGLKDQRLGQWADNITAGCGDPSKVTIFGESAGSMVLCILWNGDNT 240
DB 181 DIKAEAGSNAGLKDQRLGQWADNITAGCGDPSKVTIFGESAGSMVLCILWNGDNT 240

QY 241 YGKPLFRAGIMQSGAMVPSDPVDTGYGNEIYDLFVSSAGCGSADKLACLSASDITLL 300
DB 241 YGKPLFRAGIMQSGAMVPSDPVDTGYGNEIYDLFVSSAGCGSADKLACLSASDITLL 300

QY 301 DATNTPGFLAYSRLSLYLPRPDGKNITDDMYKLVDRGKYASVPVLIIGDQNDGTFI 360
DB 301 DATNTPGFLAYSRLSLYLPRPDGKNITDDMYKLVDRGKYASVPVLIIGDQNDGTFI 360

QY 361 SSINTTNAQARAYFKQSFHASDAEIDTLMAYPQDITQGSPPFTGVLNALTPOPKRIS 420
DB 361 SSINTTNAQARAYFKQSFHASDAEIDTLMAYPQDITQGSPPFTGVLNALTPOPKRIS 420

QY 421 AVLGDLAFTHARRYFLNHFQGGTKYSLKQSLGPIMGCTEHANDIVMODYLLGSGSVIY 480
DB 421 AVLGDLAFTHARRYFLNHFQGGTKYSLKQSLGPIMGCTEHANDIVMODYLLGSGSVIY 480

QY 481 NNAFTAFATDLDNPNTAGLLVNWPKYTSSTSSQSGNNLMINALGLYTCKONFRTAGYDALMT 540
DB 481 NNAFTAFATDLDNPNTAGLLVNWPKYTSSTSSQSGNNLMINALGLYTCKONFRTAGYDALMT 540

RESULT 3
ID ABU09073
AC ABU09073;
XX ABU09073;
XX ABU09073;
DT 15-JUL-2003 (first entry)
XX
DE Candida rugosa lipase 5 polypeptide.
XX
KW Lipase 5; enzyme; mutant; mutein.
XX
OS Candida rugosa.
XX
FH Key Location/Qualifiers
FT Misc-difference 1. .13
FT FT /note= "Absent in wild-type Candida rugosa lipase"
FT Misc-difference 14
FT FT /note= "Wild-type Ala substituted by Val"
FT Misc-difference 160
FT FT /note= "Wild-type Lys substituted by Glu"
FT Misc-difference 269
FT FT /note= "Wild-type Thr substituted by Ala"
FT Misc-difference 359
FT FT /note= "Wild-type Gly substituted by Asp"
FT Misc-difference 503
FT FT /note= "Wild-type Ser substituted by Tyr"

```

```

XX EPI288294-A2.
PN XX
PD PD
XX 05-MAR-2003.
XX
PF 26-APR-2002; 2002EP-00009616.
XX
XX 31-AUG-2001; 2001US-00943857.
PR
XX (SINI-) ACAD SINICA.
PA
XX Tang S, Lee G, Shaw J;
PI
XX WPI; 2003-395476/38.
DR
XX N-PSDB; ABX95908.
XX
PT Isolated mutant nucleic acid encoding Candida rugosa lipase, useful for
PT the preparation of Candida rugosa lipase for biocatalytic applications.
XX
PS Claim 35; Page 8-9; 33pp; English.
XX
CC The invention relates to an isolated mutant nucleic acid encoding a
CC Candida rugosa lipase polypeptide. The DNA has a sequence having at least
CC 80% identity to a wild-type DNA encoding Candida rugosa lipase. The
CC invention also relates to a microorganism comprising the DNA, where the
CC microorganism is a bacterium or yeast, preparing a mutant DNA encoding a
CC C. rugosa lipase and a chimeric C. rugosa lipase comprising a substrate
CC interacting domain of a first C. rugosa lipase and a non-substrate
CC interacting domain of a second C. rugosa lipase. The method is useful for
CC preparing a mutant DNA encoding a Candida rugosa lipase. The DNA is
CC useful in the large scale manufacture of Candida rugosa lipase which is
CC useful for biocatalytic applications. This sequence represents Candida
CC rugosa lipase 5
XX
SQ Sequence 547 AA;
Query Match      87.1%; Score 2495; DB 6; Length 547;
Best Local Similarity 86.3%; Pred. No. 2.7e-212;
Matches 472; Conservative 31; Mismatches 44; Indels 0; Gaps 0;

QY 1 SMNSRGPAGRLGSPVTAKLANGDTTGLNAINAEFLGIPFAEPVGNLRFKDPVPYSGS 60
DB 1 SMNSRGPAGRLGSPVTATLANGDTTGLNAINAEFLGIPFAEPVGNLRFKDPVPYSGS 60

QY 61 LINGQKFTSYGSCMQNPGTEENLGKTDALDLMQSKVFOAVLPQSEDLTINVRPFG 120
DB 61 LINGQKFTSYGSCMQNPGTEENLGKTDALDLMQSKVFOAVLPQSEDLTINVRPFG 120

QY 121 TKAGANLPVLMWIFGGGFEIGSPTIPPPAQMYTKSVLMGKHIIHVAVNYRVASWGFLAGD 180
DB 121 TKAGANLPVLMWIFGGGFEIGSPTIPPPAQMYTKSVLMGKHIIHVAVNYRVASWGFLAGD 180

QY 181 DIKAEAGSNAGLKDQRLGQWADNITAGCGDPSKVTIFGESAGSMVLCILWNGDNT 240
DB 181 DIKAEAGSNAGLKDQRLGQWADNITAGCGDPSKVTIFGESAGSMVLCILWNGDNT 240

QY 241 YGKPLFRAGIMQSGAMVPSDPVDTGYGNEIYDLFVSSAGCGSADKLACLSASDITLL 300
DB 241 YGKPLFRAGIMQSGAMVPSDPVDTGYGNEIYDLFVSSAGCGSADKLACLSASDITLL 300

QY 301 DATNTPGFLAYSRLSLYLPRPDGKNITDDMYKLVDRGKYASVPVLIIGDQNDGTFI 360
DB 301 DATNTPGFLAYSRLSLYLPRPDGKNITDDMYKLVDRGKYASVPVLIIGDQNDGTFI 360

QY 361 SSINTTNAQARAYFKQSFHASDAEIDTLMAYPQDITQGSPPFTGVLNALTPOPKRIS 420
DB 361 SSINTTNAQARAYFKQSFHASDAEIDTLMAYPQDITQGSPPFTGVLNALTPOPKRIS 420

QY 421 AVLGDLAFTHARRYFLNHFQGGTKYSLKQSLGPIMGCTEHANDIVMODYLLGSGSVIY 480
DB 421 AVLGDLAFTHARRYFLNHFQGGTKYSLKQSLGPIMGCTEHANDIVMODYLLGSGSVIY 480

QY 481 NNAFTAFATDLDNPNTAGLLVNWPKYTSSTSSQSGNNLMINALGLYTCKONFRTAGYDALMT 540

```

```

|||||
481 NNAFIATDLDNNTAGLSVQWPKYTSSQAGNLMQISALGLYTGKDNFRTAGYNALFA 540
|||||
541 NPSSFFV 547
: || |||
541 DPGSHFFV 547

RESULT 4
AAB47622
ID AAB47622 standard; protein; 534 AA.
XX
AC AAB47622;
XX
XX 21-JAN-2002 (first entry)
DT
XX
XX Synthetic lip1.
XX
XX Lipase; lip1; variant; ripening form; serine; 16-18C acyl chain.
XX
XX Candida rugosa.
OS
XX
XX EP1130100-A1.
XX
XX 05-SEP-2001.
XX
XX 02-FEB-2001; 2001EP-00200375.
XX
XX 14-FEB-2000; 2000EP-00200513.
XX
XX (UNIL ) UNILEVER NV.
XX
XX (UNIL ) UNILEVER PLC.
XX
XX Brocca S, Bornscheuer UT, Pleiss J, Schmid RD, Schmid U;
XX
XX Schmitt J;
XX
XX WPI; 2001-649825/75.
XX
XX N-PSDB; AAB43625.
XX
XX Modified lipolytic enzymes with altered substrate specificity, useful for
XX
XX biocatalytic applications comprising high specificity towards carbon 16
XX
XX and carbon 18 acyl chains.
XX
XX Claim 1; Fig 1; 33pp; English.
XX
XX This sequence shows a parent lipase, lip1, derived from C. rugosa. The
XX
XX lipase of the invention is a variant of this parent lipase, with altered
XX
XX properties. The variant is the ripening form of C.rugosa lipase selected
XX
XX from pre, pro, prepro or mature lipase, in which 60% or less of the CTG
XX
XX codons encoding serine in the native C.rugosa sequence, are replaced by a
XX
XX universal codon for serine. The modified nucleic acid sequence is further
XX
XX modified, such that lipase variant exhibits an altered property. The
XX
XX modified lipase is useful in a process requiring high specificity towards
XX
XX 16-18C acyl chains
XX
XX Sequence 534 AA;
XX
XX Query Match 85.9%; Score 2459; DB 4; Length 534;
XX
XX Best Local Similarity 86.7%; Pred. No. 4.1e-209;
XX
XX Matches 462; Conservative 29; Mismatches 42; Indels 0; Gaps 0;
XX
XX 15 PTAKLANGDTITGLNAINAEFLGIPFAEPVGNLRFKDPVPSGSLNGQKFTSYGSPSCM 74
XX
XX 2 PTATLANGDTITGLNAINAEFLGIPFAEPVGNLRFKDPVPSGSLNGQKFTSYGSPSCM 61
XX
XX 75 QONPEGTEENLGTALDLMVQSKVQAVLPOSEDCLTINVRPPTGAGANLPVMLWIF 134
XX
XX 62 QONPEGTEENLPLAALDLVQSKVFEAVSPSEDCLTINVRPPTGAGANLPVMLWTF 121
XX
XX 135 GGGFEIGSPTTFPPAQWTKSVLMGKHIIHVAVNVRVASWGLAGDDIKAEKSGNAGLKD 194
XX
XX 122 GGGFETGTGTRTFPPAQMITKSIAMGKPIIHVSINVRSSWGLAGDEIKAEKSGNAGLKD 181

```

```

195 QRLGMQWADNIAGFGDPSKVTIFGESAGSMVLCILWNGDNTYKGPPLFRAGIMQS 254
|||||
182 QRLGMQWADNIAAFGGDPTKVTIFGESTTSMVMCHILWNGDNTYKGPPLFRAGIMQS 241
|||||
255 GAMVPSDPVDTGTYGNEIYDLFVSSAGCGSASDKLACLRSSASDITLDDATNTTGTFLAYSS 314
|||||
242 GAMVPSDAVDGIYGNEIFDILLASNAGCGSASDKLACLRGVSSETLEDAITNTTGTFLAYSS 301
|||||
315 LRLSYLPRPDGKNITDDMYKLVDRDGKVASVPVLIIGDQNDGDTIFGLSSLVNTTNAQARAY 374
|||||
302 LRLSYLPRPDGVITDDMYALVREGKYNIPVLIIGDQNDGDTIFGLSSLVNTTNAQAREY 361
|||||
375 FKQSFHASDAEIDTLMAAYPQDITQSPEDTGVNLALTPOFKRISAVLGLAFIHARRY 434
|||||
362 FKQSFHASDTEIDTMTAVPGDITQSPEDTGVNLALTPOFKRISAVLGLGFTLARRY 421
|||||
435 FLNHFQGGTKYSLSKOLSGIPMGTFPHANDIYWDYLLGSGSVIYNNAFAFATDIDPN 494
|||||
422 FLNHYTGGTKYSLSKOLSGIPVLTGTPHNSNDIVFDYLLGSGSVIYNNAFAFATDIDPN 481
|||||
495 TAGLLVNWPKYTSSSQSGNNLMINLGLYTGKDNFRTAGYDALMTNPSSFFV 547
|||||
482 TAGLLVNWPEYTSSTSSQSGNNLMINLGLYTGKDNFRTAGYDALFENPPSFFV 534

RESULT 5
ABU09072
ID ABU09072 standard; protein; 547 AA.
XX
AC ABU09072;
XX
XX 15-JUL-2003 (first entry)
DT
XX
XX Candida rugosa lipase 4 polypeptide.
DE
XX
XX Lipase 4; enzyme; mutant; mutein.
XX
XX Candida rugosa.
OS
XX
XX Key Location/Qualifiers
XX
XX Misc-difference 1..13 "Absent in wild-type Candida rugosa lipase"
XX
XX Misc-difference 14 "Wild-type Ala substituted by Val"
XX
XX EP1288294-A2.
XX
XX 05-MAR-2003.
XX
XX 26-APR-2002; 2002EP-00009616.
XX
XX 31-AUG-2001; 2001US-00943857.
XX
XX (SINI-) ACAD SINICA.
XX
XX Tang S, Lee G, Shaw J;
XX
XX WPI; 2003-395476/38.
XX
XX N-PSDB; ABX95907.
XX
XX Isolated mutant nucleic acid encoding Candida rugosa lipase, useful for
XX
XX the preparation of Candida rugosa lipase for biocatalytic applications.
XX
XX Claim 34; Page 7-8; 33pp; English.
XX
XX The invention relates to an isolated mutant nucleic acid encoding a
XX
XX Candida rugosa lipase polypeptide. The DNA has a sequence having at
XX
XX 80% identity to a wild-type DNA encoding Candida rugosa lipase. The
XX
XX invention also relates to a microorganism comprising the DNA, where
XX
XX the microorganism is a bacterium or yeast, preparing a mutant DNA encoding a
XX
XX C.rugosa lipase and a chimeric C.rugosa lipase comprising a substrate
XX
XX interacting domain of a first C.rugosa lipase and a non-substrate
XX
XX interacting domain of a second C.rugosa lipase. The method is useful for

```

CC preparing a mutant DNA encoding a Candida rugosa lipase. The DNA is
 CC useful in the large scale manufacture of Candida rugosa lipase which is
 CC useful for biocatalytic applications. This sequence represents Candida
 CC rugosa lipase 4
 XX
 SQ Sequence 547 AA;

Query Match 85.0%; Score 2434; DB 6; Length 547;
 Best Local Similarity 83.5%; Pred. No. 7e-207;
 Matches 457; Conservative 38; Mismatches 52; Indels 0; Gaps 0;

QY 1 SMNSRGPAGRLGVPVTAFLANGDTITGLNAINAEALGIPFAEPVGNLRFKDPVPYSGS 60
 DB 1 SMNSRGPAGRLGVPVTAFLANGDTITGLNAINAEALGIPFAEPVGNLRFKDPVPYSGS 60
 QY 61 LMGKFTSYGSPSCMQNPEGTPEENLGKTDALVMSQKVFQAVLPQSEDCLTINVRPPG 120
 DB 61 LMGKFTSYGSPSCMQNPEGTPEENLGKTDALVMSQKVFQAVLPQSEDCLTINVRPPG 120
 QY 121 TKAGANLPVMLWIFGGGFEIGSPITFPFAQMVTKSVLMGKHIIHVAVNYRVASWGLAGD 180
 DB 121 TKPGANLPVWVIFGGFEVGGSSLPFAQMITASVLMGKPIIHVSMNYRVASWGLAGP 180
 QY 181 DIKAECSGNAGLKDQRLGMQWADNITAGFGGDPGSKVTIFGESAGSMVLCILWINDGNT 240
 DB 181 DIKAECSGNAGLKDQRLGLQWADNITAGFGGDPGSKVTIFGESAGSMVLCILWINDGNT 240
 QY 241 YGKPLFRAGIMQSGAMVSDPDVGYGNEIYDLFVSSAGCGSASDKLACLRASDSTLL 300
 DB 241 YGKPLFRAMIMQSGAMVSDPDVGYGQIYDQVVASAGCGSASDKLACLRASISNDKLF 300
 QY 301 DATNNTPGFLAYSSLRLSFLPRPDGKNIITDDMYKLVDRGKYASVPVVIIGDQNDGTFGL 360
 DB 301 QATSDTPGALAYPSLRLSFLPRPDGTFITDDMFKLVRDCKANVPVVIIGDQNDGTFEAL 360
 QY 361 SSLNVTNQAARAYFKQSFHASDAEIDTILMAAYPDITQGSPPFTGVLNALTPOFKRIS 420
 DB 361 SSLNVTDAQARQYFKESFIHASDABIDTILMAAYPSDITQGSPPFTGIFNAITPOFKRIA 420
 QY 421 AVIGDLAFTHARRYFLNHFQGGTKYSLKSLGSLPIMGTTHANDIVWQDYLGGSGSVIY 480
 DB 421 AVIGDLAFTHARRYFLNHFQGGTKYSLKSLGSLPIMGTTHANDIVWQDYLGGSGSVIY 480
 QY 481 NNAFTAFATDLPNTAGLVNPKYTSQSGNNLMINALGLYTCKDNFRTAGYDALMT 540
 DB 481 NNAFTAFANDLPNKAGLVNPKYTSQSGNNLQINALGLYTCKDNFRTAGYDALFT 540
 QY 541 NPSSEFV 547
 DB 541 NPSSEFV 547

RESULT 6
 ABU09070
 ID ABU09070 standard; protein; 547 AA.
 XX ABU09070;
 AC
 XX
 XX
 DT 15-JUL-2003 (first entry)
 XX
 DE Candida rugosa lipase 2 polypeptide.
 XX
 XX Lipase 2; enzyme; mutant; muten.
 XX
 XX Candida rugosa.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 1..13
 FT /note= "Absent in wild-type Candida rugosa lipase"
 FT Misc-difference 14
 FT /note= "wild-type Ala substituted by Val"
 FT Misc-difference 48
 FT /note= "wild-type Thr substituted by Ser"

FT Misc-difference 91
 FT /note= "Wild-type Arg substituted by Leu"
 FT Misc-difference 92
 FT /note= "Wild-type His substituted by Asp"
 XX
 PN EP1288294-A2.
 PD
 XX 05-MAR-2003.
 XX
 PF 26-APR-2002; 2002EP-00009616.
 XX
 XX 31-AUG-2001; 2001US-00943857.
 PR
 XX (SINI-) ACAD SINICA.
 PA
 XX Tang S, Lee G, Shaw J;
 XX WPI; 2003-395476/38.
 DR N-PSDB; ABX95905.
 DR
 XX Isolated mutant nucleic acid encoding Candida rugosa lipase, useful for
 PT the preparation of Candida rugosa lipase for biocatalytic applications.
 XX
 PS Claim 35; Page 4-5; 33pp; English.
 CC The invention relates to an isolated mutant nucleic acid encoding a
 CC Candida rugosa lipase polypeptide. The DNA has a sequence having at least
 CC 80% identity to a wild-type DNA encoding Candida rugosa lipase. The
 CC invention also relates to a microorganism comprising the DNA, where the
 CC microorganism is a bacterium or yeast, preparing a mutant DNA encoding a
 CC C. rugosa lipase and a chimeric C. rugosa lipase comprising a substrate
 CC interacting domain of a first C. rugosa lipase and a non-substrate
 CC interacting domain of a second C. rugosa lipase. The method is useful for
 CC preparing a mutant DNA encoding a Candida rugosa lipase. The DNA is
 CC useful in the large scale manufacture of Candida rugosa lipase which is
 CC useful for biocatalytic applications. This sequence represents Candida
 CC rugosa lipase 2
 XX
 SQ Sequence 547 AA;

Query Match 83.7%; Score 2396; DB 6; Length 547;
 Best Local Similarity 82.3%; Pred. No. 1.7e-203;
 Matches 450; Conservative 34; Mismatches 63; Indels 0; Gaps 0;

QY 1 SMNSRGPAGRLGVPVTAFLANGDTITGLNAINAEALGIPFAEPVGNLRFKDPVPYSGS 60
 DB 1 SMNSRGPAGRLGVPVTAFLANGDTITGLNAINAEALGIPFAEPVGNLRFKDPVPYSGS 60
 QY 61 LMGKFTSYGSPSCMQNPEGTPEENLGKTDALVMSQKVFQAVLPQSEDCLTINVRPPG 120
 DB 61 LMGKFTSYGSPSCMQNPEGTPEENLGKTDALVMSQKVFQAVLPQSEDCLTINVRPPG 120
 QY 121 TKAGANLPVMLWIFGGGFEIGSPITFPFAQMVTKSVLMGKHIIHVAVNYRVASWGLAGD 180
 DB 121 TRASAGLPVMLWIFGGGFEIGSSLPFGDQMVAKSVLMGKPIHVSMNYRVASWGLAGP 180
 QY 181 DIKAECSGNAGLKDQRLGMQWADNITAGFGGDPGSKVTIFGESAGSMVLCILWINDGNT 240
 DB 181 DIQNEGSGNAGLHDQRLGMQWADNITAGFGGDPGSKVTIFGESAGSMSTFVHLWINDGNT 240
 QY 241 YGKPLFRAGIMQSGAMVSDPDVGYGNEIYDLFVSSAGCGSASDKLACLRASDSTLL 300
 DB 241 YGKPLFRAMIMQSGAMVSDPDVGYGTEIYNQVVASAGCGSASDKLACLRASDSTLY 300
 QY 301 DATNNTPGFLAYSSLRLSFLPRPDGKNIITDDMYKLVDRGKYASVPVVIIGDQNDGTFGL 360
 DB 301 QATSDTPGALAYPSLRLSFLPRPDGTFITDDMYALVRDGYAHVFIIGDQNDGTFGL 360
 QY 361 SSLNVTNQAARAYFKQSFHASDAEIDTILMAAYPDITQGSPPFTGVLNALTPOFKRIS 420
 DB 361 SSLNVTDAQARAYFKQSFHASDAEIDTILMAAYTSDITQGSPPFTGIFNAITPOFKRIS 420
 QY 421 AVIGDLAFTHARRYFLNHFQGGTKYSLKSLGSLPIMGTTHANDIVWQDYLGGSGSVIY 480

Db 421 ALIGDLAFTLARRYFLNYOGGKYSLSKLSGLPVLGTFHGNDDIIWQDYLVGSGSVIY 480
 QY 481 NNAFIAPATDLDNPTAGLLVNWPKYTSOSSGNNLMNIALGLYTGKDNFRTAGYDALMT 540
 Db 481 NNAFIAPANDLDNPKAGLWNWPTTSSOSSGNNLMQINGLYTGKDNFRPDAYSALFS 540
 QY 541 NPSSEFFV 547
 Db 541 NPSSEFFV 547

RESULT 7
 AAR47577
 ID AAR47577 standard; protein; 563 AA.

XX AAR47577;
 AC
 DT 16-OCT-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 19-JUL-1994 (first entry)
 XX
 DE Lipase of Geotrichum candidum.
 XX Immobilisation; enzyme; cell wall; alpha agglutinin; AGA 1; FLO 1;
 KW Major cell wall protein; glycosyl-phosphatidyl-inositol;
 KW anchoring protein; alpha factor; alpha-agglutinin; invertase; inulinase;
 KW alpha-amyase; Geotrichum candidum; lipase; enzymatic process;
 KW fermentation; biodegradation; catalysis.

XX Galactomyces geotrichum.

XX WO9401567-A1.

XX 20-JAN-1994.

XX 07-JUL-1993; 93WO-EP001763.

XX 08-JUL-1992; 92EP-00202080.

XX 14-DEC-1992; 92EP-00203899.

XX (UNIL) UNILEVER PLC.

XX (UNIL) UNILEVER NV.

XX Kllis FM, Schreuder MP, Toschka H, Verrips CT;

XX WPI; 1994-035071/04.

XX N-ESDB; AAQ54020.

XX Immobilisation of enzymes to microbial cell wall - by prodn. of fusion
 PT protein of enzyme linked to anchoring protein.

XX Example 4; Page 49-52; 99pp; English.

XX The lipase is used in a method to immobilise enzymes to a microbial cell
 CC wall. The coding sequence is used in the production of a recombinant
 CC polynucleotide which comprises a structural gene encoding a protein with
 CC catalytic activity (the lipase) and at least part of a gene encoding at
 CC least the C-terminus of a protein capable of anchoring in a eukaryotic or
 CC prokaryotic cell wall. The anchoring fragment or protein is selected from
 CC alpha agglutinin, AGA 1, FLO 1, major cell wall protein of lower
 CC eukaryotes or a proteinase of lactic acid bacteria. The recombinant
 CC polynucleotide preferably also comprises a sequence encoding a signal
 CC peptide to ensure secretion of the expressed product. The signal peptide
 CC is preferably derived from glycosyl-phosphatidyl-inositol, anchoring
 CC protein, alpha factor, alpha-agglutinin, invertase or inulinase, alpha-
 CC amyase of Bacillus or proteinases of lactic acid bacteria. The host
 CC microorganism can be used for performing enzymatic processes on an
 CC industrial scale. (Updated on 25-MAR-2003 to correct PN field.) (Updated
 CC on 16-OCT-2003 to standardise OS field)

XX Sequence 563 AA;

Query Match 38.7%; Score 1107; DB 2; Length 563;
 Best Local Similarity 43.4%; Pred. No. 4.4e-89;
 Matches 240; Conservative 73; Mismatches 196; Indels 44; Gaps 9;
 QY 9 GRIGSVPTAKLNGDITITGLNAINAEFLGIPAEPPVGNLRKFDKPPVPSGLNGQKFTS 68
 Db 16 GTLAQAFTAVLNGNEVISVLEGGKVDTFKGIPTADPPVGLRFXKHPQFTGSGYQGLKAND 75
 QY 69 YGSPCQONPEGTFE-----ENLGTALDLVMQSKVFQAVLPQSEDCLTINV 115
 Db 76 FSSACMLQDFGNALFLLDKVGLGKILPDNLRLPLYDMA-----QGSVSNEDCLYNV 129
 QY 116 VRPPGTKAGANLPVLMWIFGGFEIGSPTIFPPAQMVKTVLMGKHIIHVAIVNRVASWG 175
 Db 130 FRPAGTKPDAKLPVWVIYGGAFVFGSSASYPGNGYVKESVEMQCPVVFVSVINTRTGPYG 189
 QY 176 FLAGDDIKAEKSGNAGLKDORLGMOWADNIAGFGDPSPKVTIFGESAGSMVSLCHLWN 235
 Db 190 FLGGDAITAEAGTNAGLHDQKLEWVSDNIANFGGDPDKVMIFGESAGAMSVAHQLVAY 249
 QY 236 DGNTYKGLKPLFRAGIMQSGAMVPSDPVDGTYGNEIYDLFVSSAGC---GSASDKLACL 292
 Db 250 GGDNTYNGKQLFHSAILQSGPLPYFDSTSVGPESAYSRTFAQYAGCDTSASDNDTLACL 309
 QY 293 SASDITLLDATTN-----TPGFLAYSLSRLSYLPRPDGKNITDDMYKLVRDQKYAS 343
 Db 310 SKSSDVLHSAQNSYDLKDLFLGLLPQFLGFG-----PRPDGNIIPDAAYELYSRYAK 362
 QY 344 VPVIIGQNDGEGTIFGLSLLNVTNAQARAYFKOSFIHASDAEIDTLMAYPQDITQSGP 403
 Db 363 VPYITGQDEGTILAPVAINATTPPHVKWKLKVICQASDASLDRVLSLYPGWSGSP 422
 QY 404 FDTGVNALTPOPKRISAVLGLDLAFIHARRYFLNHFGQGTKYSPFLSKQLSGL-PIMGTFH 462
 Db 423 FRIGILNALTPOPKRIAAIFTDLLFQSPRRVMLNATKDVNRWTYATQLHNLVFLGTFH 482
 QY 463 ANDIVQDYL-LGSGSVIYNNAFTAFATDLPNTAGLLVNWPKYTSOSSGNNLMNIAL 521
 Db 483 GSDLLFYVVDLGPSSA-YRRYFISFANHHDPNVGTNLQQWDMYT---DAGKEMLQIHMI 538
 QY 522 GLYTCKDNFRTAG 534
 Db 539 GNSMRTDDFRLEG 551

RESULT 8
 AAR10330
 ID AAR10330 standard; protein; 563 AA.

XX AAR10330;

XX 24-OCT-2003 (revised)

XX 25-MAR-2003 (revised)

XX 05-APR-1991 (first entry)

XX Gene product with lipase activity.

XX ATCC 34614.

XX Galactomyces geotrichum.

XX JP02299588-A.

XX 11-DEC-1990.

XX 27-MAR-1989; 89JP-00074721.

XX 27-MAR-1989; 89JP-00074721.

XX (KURK) KURITA WATER IND LTD.

XX (OSAK) OSAKA CITY.

XX WPI; 1991-027567/04.

DR	N-PSDB; AAQ10313.	
XX	Gene for coding protein with lipase activity - is prepd. from messenger-	
PT	ribonucleic acid of geo-trichum candidum ATCC 34614.	
PT		
XX	Claim 1; Fig 4; 12bp; Japanese.	
XX		
CC	The gene product may be isolated from a transformed expression sytem, and	
CC	may be enhanced with stability in heat, alkali, acid and organic solvent	
CC	by position-specific modulation. (Updated on 25-MAR-2003 to correct PA	
CC	field.) (Updated on 24-OCT-2003 to standardise OS field)	
XX		
XX	Sequence 563 AA;	
	Query Match	38.3%; Score 1097; DB 2; Length 563;
	Best Local Similarity	43.1%; Pred.No.3.4e-88;
	Matches 238; Conservative	72; Mismatches 200; Indels 42; Gaps 8;
QY	9	GRIGSVPTAKLANGDITITGLNAINAEFLGIPFAEPVGNLRKDPVPYSGSLNGQKFTS 68
DB	16	GTLAAQATVILNGNEVLSGVLEGKVDYFKGIPFADPPVGLRFXKHPQFTGSGYQGLKAND 75
QY	69	YGPSCMOQNPEGTFE-----ENLGTKALDLVMQSKVFQAVLPOSEDCLTINV 115
DB	76	FSSACMQLDPGNATSLLDKVVGLGKIIPDNLRLGPLYDMA-----QGSVMNEDCLVNV 129
QY	116	VRPPGTKAGANLPVWLITFGGPEIGSPITFPFPAQMTKSVLMGKHIIHVAIVYRVASWG 175
DB	130	FRAGTKPDAKLPVWVIYIGAFVFGSSASYPGNAYVKESLEMQGPVVFVSINTRTGPYG 189
QY	176	FLAGDDTIKAGSGNAGLKDQRLGMQVADNIAGFGDPKVTIFGESAGSMVLCHLIWN 235
DB	190	FLCGDAITAEQNTNAGLHDQKGLEWVSDNIANFEGGDPKVMIFGESAGAMVAHQLVAY 249
QY	236	DGNTYKKGKPLFRAGIMQSGAMVPSPVDGTGYGNEIYDLFVSSAGC--GSASDK--LACL 292
DB	250	GGDNTYNGKQLFSAILOSGPLPYFDSTVGPESAYSRFQAQYAGCDASAGDNETLACL 309
QY	293	SASSDTLLDATN-----TPGFLAYSRLSYLPRDPGKNITDDMYKLVRDQKVAS 343
DB	310	SKSSDVLHSAQNSYDLKDLFLGLPQFLGFG-----PRDGNITPDAAYELRSRYAK 362
QY	344	VPVIQDQDEGTIFGLSSLVNTTNAQAARYFKOSFIHASDAEIDTLMAAYPQDITGSP 403
DB	363	VPYITGNQDEGTILAPVAINATTPHVKKWLKYICSEASDASLDRLVSLYPGWSSEGAP 422
QY	404	FDTGVLNALTPOKRIASAVLGDLAFIHARYFELNHFGGTYKYSFLSKQLSGL--PIMGTEH 462
DB	423	FRGILNALTPOKRIANAITFDLLFQSPRRVMLNATKDVNRWITYLATQLHNLPFLGTEH 482
QY	463	ANDIVMQDYLKSGSVIYNNAFIATAFDLDPNTAGLIVNWPVKYTSSSQSGNNLMINALG 522
DB	483	GSDLLFQYAGPWSSSAYRRYFISFANHDPNVAATNLKQMDWYT---DSGKEMLQIHMG 539
QY	523	LYTGKCNFRTAG 534
DB	540	NSMRTDDEFRIEG 551
	RESULT 9	
	AAR06370	
ID	AAR06370 standard; protein; 557 AA.	
XX		
AC	AAR06370;	
XX		
XX	24-OCT-2003 (revised)	
DT	25-MAR-2003 (revised)	
DT	13-DEC-1990 (first entry)	
XX		
DE	Protein with lipase activity deduced from gene.	
XX		
XX	Lipase.	
XX		

AAW39078;
 08-APR-1998 (first entry)
 Torpedo californica acetylcholinesterase AChE.
 Acetylcholinesterase; AChE; organophosphate; detoxification; oxime;
 ray: chemical agent; treatment; prevention; aging; mutant.
 Torpedo californica.
 Key Location/Qualifiers
 Modified-site 59..61
 /note= "site as indicated in specification"
 Region 67..94
 /note= "region as indicated in specification."
 Region 254..265
 /note= "region as indicated in specification"
 Region 402..521
 /note= "region as indicated in specification"
 Modified-site 416..418
 /note= "site as indicated in specification"
 Modified-site 457..459
 /note= "site as indicated in specification"
 Modified-site 569..571
 /note= "site as indicated in specification"
 US5695750-A.
 09-DEC-1997.
 25-NOV-1994; 94US-00348920.
 25-NOV-1994; 94US-00348920.
 (USSA) US SEC OF ARMY.
 Radic Z, Taylor P, Maxwell D, Doctor BP, Saxena A;
 WPI; 1998-041233/04.
 Mutant Torpedo acetylcholinesterase - useful for detoxifying organo-
 phosphate compounds.
 Disclosure; Fig 1; 12pp; English.
 This sequence represents an acetylcholinesterase (AChE) from Torpedo
 californica which is used to create enzymes capable of detoxifying
 organophosphates. Mutant versions of this AChE are useful when combined
 with an oxime for detoxifying chemical warfare agents such as sarin or as
 a means of treating or preventing the deleterious effects of
 organophosphate exposure in organisms. Such variants can also destroy
 organophosphates in an environmentally friendly manner and be more
 resistant to aging than wild-type enzymes
 Sequence 575 AA;
 Query Match 17.2%; Score 491.5; DB 2; Length 575;
 Best Local Similarity 30.4%; Pred. No. 1.9e-34;
 Matches 164; Conservative 75; Mismatches 195; Indels 105; Gaps 24;
 22 GDTITGLNAINAEPLGIPFAEPVGNLRFKDP---VPYSGSLNGQKFTSYGPGSQMGNP 78
 17 GTRVPVLSHSLS-AFLGIPFAEPVGNLRFKDP---VPYSGSLNGQKFTSYGPGSQMGNP 71
 79 EGTEENLGKTLALVMSQVFAVLPOSECLINV---VRPGTKAGANLPMWLWIF 134
 72 DEQPPFGSG-----SEMNPNREMSDECLYINIWVPSRPKST-----ITVMWIY 116
 135 GGGFEIGSPPTIPPAQMTKSLMGKH-----IIHVAVNYRVASWGFLA--GDDIKAE 186
 117 GGGPYSGSSTL-----DYNGKILAYTEEVVLVSLSYRVAGFALHGSQ---EA 164

QY 187 SGNAGLKQRLGMQWVADNIAGGGDPKVTIIFGESAGSMVLCHLIWNGDNTYKGP 246
 DB 165 PGNVGLLDQRMALQWHDNIQFFGGDPKVTIIFGESAGSAGVGMHIL-----SPGSRDL 218
 QY 247 FRAGIMQSGAMVPSDP---VDGTYGNEIYDLFVSSAGCGSADK--LACLRSSSDTLDD 301
 DB 219 FRRAILQSGS--PNCPMASVSAEGRRAVELGRNLNCNLNSDEELHCLREKKPQELID 276
 QY 302 ATNNTPGFLAYSSL-ELSYLPRPDKNITDDMKLVDRGKYASVPVLIIGQNDGTTIFGL 360
 DB 277 VEWNV---LPFDSIFRFSFVPVVDGFFFTSLSEMLNSGNFKKTQILLGVNKDEGSFLL 333
 QY 361 -----SSLNVTYNAQARAYKQSFHSDAEIDTLMAAYQDITQGSFFDTGVNLAL 412
 DB 334 YGAPGFSKDSKISREDFMSGVKLSVPHANDLGLDAVTLQY---TDMDDNNGIKNR- 388
 QY 413 TPOFKRISAVLGDLPATHARRYFLN---HFQGTGKYSFLSKQLSGL---PIMGTFHANDI 466
 DB 389 ----DGLDDIVGDHNVICPLMHFNKVTYFNGNGTYLYFFNHRASNLVWPEWGMVHGYEI 444
 QY 467 VMODYLLGSGSVIYNN-----AFIAFATDLDNTA-GLLVNWPKYTSSSQ 510
 DB 445 ---EFVGLPLVKELVNTAEELSRRIHMYWATFAKTGNPNPHSQESKWLFTTKEQ 500
 RESULT 14
 AAW39079
 ID AAW39079 standard; protein; 575 AA.
 XX AC AAW39079;
 XX DT 08-APR-1998 (first entry)
 XX DE Torpedo californica acetylcholinesterase AChE mutant E199Q.
 XX KW Acetylcholinesterase; AChE; organophosphate; detoxification; oxime;
 XX KW ray: chemical agent; treatment; prevention; aging; mutant.
 XX OS Torpedo californica.
 XX OS Synthetic.
 XX FH Key Location/Qualifiers
 FT Modified-site 59..61
 FT /note= "site as indicated in specification"
 FT Region 67..94
 FT /note= "region as indicated in specification"
 FT Modified-site 199
 FT /label= E199Q
 FT /note= "wild-type Glu is replaced by Gln"
 FT Region 254..265
 FT /note= "region as given in specification"
 FT Region 402..521
 FT /note= "region as given in specification"
 FT Modified-site 416..418
 FT /note= "site as given in specification"
 FT Modified-site 457..459
 FT /note= "site as given in specification"
 FT Modified-site 533..535
 FT /note= "site as given in specification"
 PN US5695750-A.
 XX 09-DEC-1997.
 XX 25-NOV-1994; 94US-00348920.
 XX 25-NOV-1994; 94US-00348920.
 XX (USSA) US SEC OF ARMY.
 XX Radic Z, Taylor P, Maxwell D, Doctor BP, Saxena A;
 XX WPI; 1998-041233/04.
 DR

XX Mutant Torpedo acetylcholinesterase - useful for detoxifying organo-
PT phosphate compounds.
PS Claim 1; Fig 2; 12pp; English.
XX This sequence represents a variant of the enzyme acetylcholinesterase
CC (AChE) from Torpedo californica which is capable of detoxifying
CC organophosphates. This E199Q mutant has greater resistance to aging than
CC the wild-type AChE and can destroy organophosphates in an environmentally
CC friendly manner. When combined with an oxime, this mutant's ability to
CC detoxify several organophosphates is amplified and is predicted to have
CC applications for the detoxification of chemical warfare agents such as
CC sarin or as a means of treating or preventing the deleterious effects of
CC organophosphate exposure in organisms
XX
XX Sequence 575 AA;
XX
XX Query Match 17.1%; Score 488.5; DB 2; Length 575;
XX Best Local Similarity 30.2%; Pred. No. 3.6e-34;
XX Matches 163; Conservative 76; Mismatches 195; Indels 105; Gaps 24;
XX
XX QY 22 GDTITGLNAIFAGIPFAEPVGNLRFKDP---VPYSGSLNGQKFTSYGSPCMQNP 78
XX Db 17 GTRVPVLSHS-:AFIGIFAEPPVGNMRFRPEPKPWSGVNASTY----PNNCQQYV 71
XX QY 79 EGTFEENLKTALDIAVMQSKVQAVLPQSEDCITINV---VRPPGCTKAGANLPVMLWTF 134
XX Db 72 DQFPFGSG-----SEMNPNREMSDCLYNIWVSPRPKST-----TVMWYIY 116
XX QY 135 GGGPEIGSPTIPPAQMTKSVLMGKH-----IIHVAVNRYVASWGFLA--GDDIKAE 186
XX Db 117 GGGFYSGSSTL-----DYNKGKYLAYTEEVVLSLVRVGAFFLALHGSQ---EA 164
XX QY 187 SQNAGLKQDRLGMQWADNIAFGGDPKSVTIFFGESAGSMVLCILWNGDNTYKGP 246
XX Db 165 PGNVGLLDQRMALQWHDNIQFFGGDKTIVTIFGQAGGASVGMHIL-----SPGSRDL 218
XX QY 247 FRAGIMQSGAMVPSDP---VDGTGNEIYDLFVSSAGCGSASDK--LACLRGASSDTLLD 301
XX Db 219 FRRLALQSGS--PNCFWASVVAEGRRAVELGRNLNLCNLSDEELHCLREKKPOELLID 276
XX QY 302 ATNNTPGFLAYSSL-RLSYLPRPDGKNIITDDMYKLVDRGKYASVPVILIGDQNDGTFGL 360
XX Db 277 VEWNV---LPFDSIFRSPVPIDGFEFTSLEMLNSGNFKKTKQILLGVNKGDEGSFFLL 333
XX QY 361 -----SSLVNTTNAQARAYFKQSFHISDAEIDTLMAAYPDITQGSPTDGTGLNAL 412
XX Db 334 YGAPGFSKDSKISREDPMGKUSVPHANDGLDAVTLQY---TDMMDNNGIKNR- 388
XX QY 413 TPQFKRISAVLGLAFIHARRYFLN---HFQGGTKYSLFKSLQSLG---PIMGTFHANI 466
XX Db 389 ---DGLDDIVGDHNVICPLMHFVNKYTKFGNGTYLYFFNHRASNLVWPMWGVHIGVEI 444
XX QY 467 VQDYLGLSGSVIYNN-----AFIATDLDNPTA-GLLVNWKYKTSYSSQ 510
XX Db 445 ---BFVFGPLVKELNYTAEEALSRRIHMYWATEAKTGNPNRPHSQESKSWPLFTTKEQ 500
XX
XX RESULT 15
XX ABR84598
XX ID ABR84598 standard; protein; 576 AA.
XX AC ABR84598;
XX DT 18-DEC-2003 (first entry)
XX DE T californica acetylcholinesterase protein.
XX KW Alpha-carboxylesterase; lipase; esterase; insect; biotransformation; ray;
XX biocatalysis; enzyme; hydrolysis; acetylcholinesterase.
XX OS Torpedo californica.

XX WO2003066873-A1.
XX 14-AUG-2003.
XX 06-FEB-2002; 2002WO-AU000113.
XX 06-FEB-2002; 2002WO-AU000113.
XX (CSIR) COMMONWEALTH SCI & IND RES ORG.
XX Oakeshott JG, Devonshire A, Coppin CW, Heidari R, Dorrian SJ;
XX Russell RJ;
XX WPI; 2003-767260/72.
XX Enzymatic reactions using an insect esterase or lipase, useful
XX particularly for resolution of hydrophobic insecticidal esters and
XX modification of fats and oils.
XX Example 1; Page 77-79; 82pp; English.
XX The present invention relates to an enzyme-based biocatalytic process
XX where the enzyme is an insect esterase or lipase, or its mutant. Such an
XX enzyme may be an alpha-carboxylesterase. The method is used for
XX stereospecific hydrolysis of fatty acid and insecticidal (especially
XX pyrethrin) esters, forming optically pure acids or alcohols, for
XX modification of vegetable oils or fats for use in emulsions or other fat-
XX based foods, for preparing polymers, especially polyesters, for acylation
XX of substrates in leather tanning, in detergents, for removing pitch from
XX paper pulp, in biosensors (determination of triacylglycerols) and for the
XX preparation of regio- or chirally pure pharmaceutical or agrochemical
XX products. The present sequence is the ray acetylcholinesterase enzyme,
XX which was used in a homology comparison with a protein of the invention

Sequence 576 AA;

Query Match 16.9%; Score 484; DB 7; Length 576;
Best Local Similarity 30.6%; Pred. No. 9e-34;
Matches 160; Conservative 73; Mismatches 192; Indels 98; Gaps 23;
QY 35 AFLGIPFAEPVGNLRFKDP---VPYSGSLNGQKFTSYGSPCMQNPGETFEENLKTAL 91
Db 30 AFLGIPFAEPVGNMRFRPEPKPWSGVNASTY----PNNCQQYVDQFPFGP 81
QY 92 DLVMSQKVFQAVLPQSEDCITINV-VRPPGCTKAGANLPVMLWIFGGGFEIGSPTIPPAQ 150
Db 82 -----SEMNPNREMSDCLYNIWVSPRPKSTAT---VMLWYGGFYSGSSTL----- 128
QY 151 MVTKSLVMGKH-----IIHVAVNRYVASWGFLA--GDDIKAEAGSGNAGLKQDRLGMQW 202
Db 129 ---DYNKGKYLAYTEEVVLSLVRVGAFFLALHGSQ---EAPGNWGLLDQRMALQW 181
QY 203 ADNIAGFGDPKSVTIFFGESAGSMVLCILWNGDNTYKGPFRAGIMQSGAMVPSDP 262
Db 182 HDNIQFFGDPKSVTIFFGESAGSRASVGMHIL-----SPGSRDLFRRALQSGS--PNC 233
QY 263 ---VDGTGNEIYDLFVSSAGCGSASDK--LACLRGASSDTLLDNTNTPGFLAYSSL-R 316
Db 234 WASVSAEGRRAVELRNLNLCNLSDEDLTQCLREKKPOELLIDVEWNV---LPFDSIFR 290
QY 317 LSYLPRPDGKNIITDDMYKLVDRGKYASVPVILIGDQNDGTFGL-----SSLVNTTN 368
Db 291 PSFVVIDGEFTSLEMLNAGNFKKTKQILLGVNKGDEGSFFLYGAPGFSKDSKISR 350
QY 369 AQARAYFKQSFHISDAEIDTLMAAYPDITQGSPTDGTGLNALTPQFKRISAVLGLDAF 428
Db 351 EDFMSGVKLSVPHANDGLDAVTLQY---TDMMDNNGIKNR-----DGLDDIVGNHV 401
QY 429 IHARRYFLN---HFQGGTKYSLFKSLQSLG---PIMGTFHANDIVWQDYLGLSGSVIYNN 482
Db 402 ICPLMHFVNKYTKFGNGTYLYFFNHRASNLVWPMWGVHIGVEI---BFVFGPLVKELN 458

Qy 483 -----AFIAFATDLDPNTA-GLLVNWPXYTSSQ 510
Db 459 YTAEEELSRIMHYWATFAKTGNENEPHSESKWPLFTTKEQ 501

Search completed: July 29, 2004, 10:49:10
Job time : 58 secs


```

QY 75 QONPEGTFFE-----ENLGTALDLVMQSVFOAVLPQSEDCLTINVRPPT 121
DB 63 QLDGFGNAISLLDKVVGKILPDNLGRPLYDMA-----QGSVSMNEDCLYLNVFRPAGT 116
QY 122 KAGANLPVLMWIFGGPGEIGSPITFFPAQMVTKSVLMGKHIHVAVNRYRVASWGLAGDD 181
DB 117 KPDAKLPVWMIYGGAFVFGSSASYPGNGYKESVEMQPVVFSINRYTGPYGLGDA 176
QY 182 IKAEGSGNAGLKQORLGMQWVADNIAGFGDPKSKVTTFGSGAGMSVYLCHLIWNDDGNTY 241
DB 177 ITAEGSTNAGLHDQKLGLEWSDNIANFGGDPDKVMIFGESAGAMSAHQLVAYGGDNTY 236
QY 242 KGKPLFRAGIMQSGAMVPSDPVDTGYNEIYDLFVSSAGC---GSASDKLACLSASDDT 298
DB 237 NGKQLFHSAILQSGGPLPYFDSTSVGPESAYSRAQYAGCDTSASDNDTTLACLRSKSDV 296
QY 299 LLDATNN-----TPGFLAYSLSRLSVLPRDPGKNIITDDMYKLVRDGKYASVPVITG 349
DB 297 LHSQNSYDLKDLFGLLPQFLGFG-----PRPDGNIIPDAAYELYSRGYAKVPVITG 349
QY 350 QONDEGTIFGLSSLVNTTNAQARAYFKQSFHASDAEIDTLMAYPODITQGSFDTGVL 409
DB 350 NQDEGTILAPVAINATTPHVKKWLKYICQASDASLDRLVSLYPGWSGSEGPFTGIL 409
QY 410 NALTPQKRIISAVLGDIAFHARRYFLNHFGGTYKYSFLSKQLSGL-PIMGTETHANDIVM 468
DB 410 NALTPQKRIIAAIFTDLLFQSPRRVMLNATKDVNRWYIATQLHNLVPLFTFGHSDLLF 469
QY 469 QDYL-LGSGSVIYNNAFIATATDLPNTAGLLVNNPKYTSQSGNNLMINALGLYTKG 527
DB 470 QYVVDLGPSSA-YRRYFISFANHDPNVGTLNQWDMYT---DAGKEMLIHMGNSMRT 525
QY 528 DNFRTAG 534
DB 526 DDFRIEG 532

RESULT 2
Q00882 PRELIMINARY; PRT; 544 AA.
AC Q00882
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Lipase (Fragment).
OS Geotrichum candidum (Oospora lactis).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Galactomyces.
OX NCBI_TaxID=27317;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-553;
RA Bertolini M., Laramee L., Thomas D., Cygler M., Schrag J., Vernet T.;
RT "Polymorphism in the lipase genes of Geotrichum candidum strains.";
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLSTERASE/LIPASE FAMILY.
DR EMBL; U02387; AAA03425.1; -.
DR PIR; S41093; S41093.
DR HSSP; P17573; 1THG.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR002018; Carboxylesterase.
DR InterPro; IPR000379; Ser_estrs.
DR Pfam; PF00135; Coesterase; 1.
DR PROSITE; PS00122; CARBOXYLSTERASE_B_1; 1.
DR PROSITE; PS00941; CARBOXYLSTERASE_B_2; 1.
KW Hydrolase.
FT NON_TER 1
SQ SEQUENCE 544 AA; 59264 MW; 96002E6652C23593 CRC64;

Query Match 38.0%; Score 1089; DB 3; Length 544;
Best Local Similarity 43.5%; Pred. No. 9,1e-68;
Matches 238; Conservative 73; Mismatches 192; Indels 44; Gaps 10;

```

```

QY 15 PTKLANGDTITGLNAILINEAFLGIPPAEPPVGNLRFKDPVPVYSGSLNGKQFTSYGSCM 74
DB 3 PTVANGNEVISGLVGKVDTFKGIFFADPPVGLRFPKHQPPTGTSYQGLKANDFSACM 62
QY 75 QONPEGTFFE-----ENLGTALDLVMQSVFOAVLPQSEDCLTINVRPPT 121
DB 63 QLDGFGNAISLLDKVVGKILPDNLGRPLYDMA-----QGSVSMNEDCLYLNVFRPAGT 116
QY 122 KAGANLPVLMWIFGGPGEIGSPITFFPAQMVTKSVLMGKHIHVAVNRYRVASWGLAGDD 181
DB 117 KPDAKLPVWMIYGGAFVFGSSASYPGNGYKESVEMQPVVFSINRYTGPYGLGDA 176
QY 182 IKAEGSGNAGLKQORLGMQWVADNIAGFGDPKSKVTTFGSGAGMSVYLCHLIWNDDGNTY 241
DB 177 ITAEGSTNAGLHDQKLGLEWSDNIANFGGDPDKVMIFGESAGAMSAHQLVAYGGDNTY 236
QY 242 KGKPLFRAGIMQSGAMVPSDPVDTGYNEIYDLFVSSAGC---GSASDKLACLSASDDT 298
DB 237 NGKQLFHSAILQSGGPLPYFDSTSVGPESAYSRAQYAGCDTSASDNDTTLACLRSKSDV 296
QY 299 LLDATNN-----TPGFLAYSLSRLSVLPRDPGKNIITDDMYKLVRDGKYASVPVITG 349
DB 297 LHSQNSYDLKDLFGLLPQFLGFG-----PRPDGNIIPDAAYELYSRGYAKVPVITG 349
QY 350 QONDEGTIFGLSSLVNTTNAQARAYFKQSFHASDAEIDTLMAYPODITQGSFDTGVL 409
DB 350 NQDEGTILAPVAINATTPHVKKWLKYICSEASDASLDRLVSLYPGWSGSEGPFTGIL 409
QY 410 NALTPQKRIISAVLGDIAFHARRYFLNHFGGTYKYSFLSKQLSGL-PIMGTETHANDIVM 468
DB 410 NALTPQKRIIAAIFTDLLFQSPRRVMLNATKDVNRWYIATQLHNLVPLFTFGHSDLLF 469
QY 469 QDYL-LGSGSVIYNNAFIATATDLPNTAGLLVNNPKYTSQSGNNLMINALGLYTKG 527
DB 470 QYVVDLGPSSA-YRRYFISFANHDPNVGTLNQWDMYT---DAGKEMLIHMGNSMRT 525
QY 528 DNFRTAG 534
DB 526 DDFRIEG 532

RESULT 3
Q00886 PRELIMINARY; PRT; 544 AA.
AC Q00886
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Lipase (Fragment).
OS Geotrichum candidum (Oospora lactis).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Galactomyces.
OX NCBI_TaxID=27317;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 34614;
RA Bertolini M., Laramee L., Thomas D., Cygler M., Schrag J., Vernet T.;
RT "Polymorphism in the lipase genes of Geotrichum candidum strains.";
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLSTERASE/LIPASE FAMILY.
DR EMBL; U02622; AAA03435.1; -.
DR PIR; S41090; S41090.
DR HSSP; P17573; 1THG.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR002018; Carboxylesterase.
DR InterPro; IPR000379; Ser_estrs.
DR Pfam; PF00135; Coesterase; 1.
DR PROSITE; PS00122; CARBOXYLSTERASE_B_1; 1.
DR PROSITE; PS00941; CARBOXYLSTERASE_B_2; 1.
KW Hydrolase.
FT NON_TER 1
SQ SEQUENCE 544 AA; 59263 MW; 97C96E7807DBF1EC CRC64;

```

```

Query Match      38.0%; Score 1088; DB 3; Length 544;
Best Local Similarity 43.5%; Pred. No. 1.1e-67;
Matches 238; Conservative 72; Mismatches 193; Indels 44; Gaps 10;

QY 15 PTAKLANGTTITGLNAINEAFGLIPFAEPVGNLRFKDPVYSGSLNGOKFTSYGSPSCM 74
D 3 PTAVLNGNEISGVLEKGVDTFKGIPFADPVGDLRFKHPQPTFGSYQGLKANDFFSACM 52
QY 75 QONPECTFE-----ENLKTALDLVMQSKVQAVLPQSEDCLTINVRPPGT 121
D 63 QLDPGNAISLLDKVGLGIIPDLNGLPLYDMA-----QGSVMNEDCLILNVFRPAGT 116
QY 122 KAGANLPVMLTIFGGGFEIGSPITFPFAQMVTKSVLMGXHIIHVAVNRYRVASGFLAGDD 181
D 117 KPDAKLPVMMWIIYGGAFVFGSSASYPGNGYVXESVEMGQPVVFSINRYRTPGYFLGGDA 176
QY 182 IKAEGSGNAGLDORLGMQWADNIAFGDGPDKVTFIAGESAGSMVSLCHLIWDGDNITY 241
D 177 ITAEGNTNAGLHDQRKLEWSDNIAFGDGPDKVMIFGESAGAMVAHQIA 236
QY 242 KGKPLFRAGIMOSGAMVSDPDVDTYGNIEYDLFVSSAGC-GSADK--LACLRSSASDT 298
D 237 NGKQLFHSAILQSGPLPYFDSTSVGPESAYSREFAQYAGCDASAGNETLACLRSSSDV 296
QY 299 ILDATNN-----TPGLAYSRLSYLPRPDGKNITDDMYKLVDRGKYASVPIIG 349
D 297 LHSAGNSYDLKDLFGLLPQFLGFG-----PRPDGNIIPDAAYELRYGRYAKVPIITG 349
QY 350 DONDEGTIFGLSLLNVTTNAQARAYFKOSFIHASDAEIDLMAAYPQDITQSPPTGVL 409
D 350 NQDEGTILAPVAINATTTPHVKWLVKTYTCSEASDASLDRLVLSLXPGWSEGAFFRTGIL 409
QY 410 NALTPOFKRISAVLGLDLAFIARRYFLNHFQGGTKYSLSKLSGL-PIMGTFHANDIYV 468
D 410 NALTPOFKRIALTDLQSFRRVMLNATKQVNRWYTLATQHLNVLPFLGTFHSGDLIF 469
QY 469 QYLL-LGSGSVIYNNAFIATDLDNTAGLLVNWPKYTSOSSGNNLMNIALGLYTKG 527
D 470 QYVVDLGPSSA-YRRYFISFANHDPNVGTNLKQWDMYT---DSGKEMIQHIMGNMST 525
QY 528 DNFRTAG 534
D 526 DDFRIEG 532

RESULT 4
ID P79066 PRELIMINARY; PRT; 563 AA.
AC P79066;
DT 01-MAY-1997 (TremBLrel. 03, Created)
DT 01-MAY-1997 (TremBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Lipase I precursor.
OS Geotrichum fermentans.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; mitosporic Dipodascaceae;
OC Geotrichum
OX NCBI_TaxID=44066;
RN [1]_TaxID=44066;
RP SEQUENCE FROM N.A.
RC STRAIN=WU-C12;
RA Arai T.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-188 FROM N.A.
RC STRAIN=WU-C12;
RX MEDLINE=97372566; PubMed=9228786;
RA Arai T., Yusa S., Kikimura K., Usami S.;
RT "Cloning and sequencing of the pcDNA encoding lipase I from
RT Trichosporon fermentans WU-C12."
RL PEMS Microbiol. Lett. 152:183-188 (1997).
CC -!- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.

```

```

DR EMBL; AB000260; BAA19072.1; -.
DR HSSP; P17573; 1THG.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR002018; CarboxylaseB.
DR InterPro; IPR000379; Ser_estrs.
DR Pfam; PF00135; Coesterase; 1.
DR PROSITE; PS00122; CARBOXYLESTERASE_B_1; 1.
DR PROSITE; PS00941; CARBOXYLESTERASE_B_2; 1.
KW Hydrolase; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 563 POTENTIAL.
SQ SEQUENCE 563 AA; 61548 MW; 210A12C206F33881 CRC64;

Query Match      38.0%; Score 1087; DB 3; Length 563;
Best Local Similarity 42.0%; Pred. No. 1.3e-67;
Matches 235; Conservative 83; Mismatches 200; Indels 42; Gaps 8;

QY 8 AGRLSVPTAKLANGTTITGLNAINEAFGLIPFAEPVGNLRFKDPVYSGSLNGOKFT 67
D 15 AGVLAQAPTAVLNGNEISGVLEKGVDTFKGIPFADPVGDLRFKHPQPTFGSYQGLKAN 74
QY 68 SYGPSQWQONPEGTPEENLKTALDLVM-QSKVF-----QAVLPOSEDCITIN 114
D 75 DFSPACMQJDPGNSL-----TLDDKALGLAKVIPPEFRGPLYDMAKGTVMNEDCLYN 128
QY 115 VVRPPTKAGANLPVMLTIFGGGFEIGSPITFPFAQMVTKSVLMGXHIIHVAVNRYVASV 174
D 129 VFRPAGTKPDAKLPVMMWIIYGGAFVYGSSAAYPGNSYVYESINMGQPVVFSINRYTGP 188
QY 175 GFLAGDDIKAGSGNAGLDORLGMQWADNIAFGDGPDKVTFIAGESAGSMVSLCHLIW 234
D 189 GFLGGDAITAEAGNTNAGLHDQRKLEWSDNIAFGDGPDKVMIFGESAGAMVAHQIA 248
QY 235 NQDNTYKGLPIFRAGIMOSGAMVSDPDVDTYGNIEYDLFVSSAGC---GSADKLA 291
D 249 YGDMNTYNGKLFHSAIQLQSGPLPYHDSVSSVGPDIISYNRFAQYAGCDTSASANDTLE 308
QY 292 RGASDSTLLDATNN-----TPGLAYSRLSYLPRPDGKNITDDMYKLVDRGKYA 342
D 309 RKSSSLVLDQAQSYDLKDLFGLLPQFLGFG-----PRPDGNIIPDAAYELFRSGRYA 361
QY 343 SVPIVIGQNDGTIFGLSLLNVTTNAQARAYFKOSFIHASDAEIDLMAAYPQDITQGS 402
D 362 KVPYISGNQDEGTAFAPVALNATTTPHVKWLVQIFVDASEASIDRLVLSLPQTLVSGS 421
QY 403 PFDTCVLANALTPOFKRISAVLGLDLAFIARRYFLNHFQGGTKYSLSKLSGL-PIMGTF 461
D 422 PPRTGILNALTPOFKRVAAILSDMLFQSPRRVMLSATKDQVNRWYTLSTHLNLPFLGTF 481
QY 462 HANDIYVQDYLLGSGSVIYNNAFIATDLDNTAGLLVNWPKYTSOSSGNNLMNIAL 521
D 482 HGNELIFQFNVIQFNVIQFNVIQFNVIQFNVIQFNVIQFNVIQFNVIQFNVIQFNVI 538
QY 522 GLYTKGNFRAGYDALMTN 541
D 539 DNVMRTDDYRIEGISNFFET 558

RESULT 5
ID Q00883 PRELIMINARY; PRT; 544 AA.
AC Q00883;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Lipase (Fragment).
OS Geotrichum candidum (Oospora lactis).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Galactomycetes.
OX NCBI_TaxID=27317;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-552;

```

```
RA Bertolini M., Laramee L., Thomas D., Cygler M., Schrag J., Vernet T.;  
RT "Polymorphism in the lipase genes of Geotrichum candidum strains."  
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.  
DR EMBL; U02524; AAA03428.1; --  
DR PIR; S41092; S41092.  
DR HSP; P17573; ITHG.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR002018; CarboxylesteraseB.  
DR Pfam; PF00135; Coesterase; 1.  
DR PROSITE; PS00122; CARBOXYLESTERASE_B_1; 1.  
DR PROSITE; PS00941; CARBOXYLESTERASE_B_2; 1.  
KW Hydrolase.  
FT NON_TER  
SQ SEQUENCE 544 AA; 59257 MW; 85DB41B1E28B8BF1 CRC64;  
  
Query Match 37.8%; Score 1084; DB 3; Length 544;  
Best Local Similarity 43.3%; Pred. No. 2e-67;  
Matches 237; Conservative 73; Mismatches 193; Indels 44; Gaps 10;  
  
QY 15 PTKLANGDITITGLNAINEAFGLIPAEPPVGNLRKDPVPYSGSLNGKFTSYGSPSCM 74  
DB 3 PTVANGNEVISGLVGKVDTFKGIIPADPPVGLRFPKFPPTGSGYQGLKANDFSSACM 62  
QY 75 QONPEGTFFE-----ENLGTALDLVMQSKVFOAVLPQSEDCLTINVRPPT 121  
DB 63 QLDPGNAISLLDKVVGKGIIPDLNLRGLPYDMA-----QGSVMNEDCLYLVNFRPAGT 116  
QY 122 KAGANLPMVLWIFGGFEGISPTIFFPPAQMTKSLVLMKGIHIVAVNRYVASWGFLAGDD 181  
DB 117 KPDAKLPVMWIIYGAFFVGSSASYPGNGYKESVEMGQPVVFSINVRTGPGYFLGGDA 176  
QY 182 IKAGSGNAGLKQORLQWVADNIAGFGDPSKVTIFGESAGSMVSLCHLIWNGDNTY 241  
DB 177 ITAEGNTNAGLHDQKGLWVSDNIANFGGDPDKVMIFGESAGAMSAHQIIVAYGGDNTY 236  
QY 242 KGKPLFRAGIMQSGAMVPSDPVDGTGNEYIYDLVFSSAGC--GSASDK--LACLSASD 298  
DB 237 NGKQLFHSAILQSGPLPYFSTSVGPESASVSRFAQYAGCDTSVSDNDTLACLRSKSDV 296  
QY 299 LLDATNN-----TPGFLAYSSLRLSYLPRDGKNITDDMYKLVDRDKYASVPVITG 349  
DB 297 LHSQNSYDLKDLFGLLPQFLGFG-----PRPDGNIIPDAAYELYSRGYAKVPYITG 349  
QY 350 DONDEGTIFGLSSLVNTTNAQARAYFKQSFTHASDAEIDTLMAAYPQDITQSPFDITGV 409  
DB 350 NODEGTILAPVAINATTPHVKWLYICSEASDAPLDRVLSLYPGSWSEGAPFRITGIL 409  
QY 410 NALTPQFKRIASVLDGLAFIHARRYFLNHFGGKTSFLSKQLSGL--PIMGTFHNDITW 468  
DB 469 QDYL--LGSGSVIYNNAFAFATDLPNTAGLLVNNPKYTSQSGNNLMNINAGLYTGK 527  
DB 470 QYYVDLGFSSA--YRRYFISFANHDPNVGTNLKQWDMYT---DAGKEMLIQIHMGNSMRT 525  
QY 528 DNFRTAG 534  
DB 526 DDFRIEG 532  
  
RESULT 6  
Q12614 PRELIMINARY; PRT; 544 AA.  
AC Q12614;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
DE Triacylglycerol lipase (EC 3.1.1.3) (Fragment).  
OS Geotrichum candidum (Oospora lactis).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Dipodascaceae; Galactomycetes.
```

```
OX NCBI_TaxID=27317;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CBS 178.71;  
RX MEDLINE=96049515; PubMed=7578238;  
RA Phillips A., Pretorius G.H.J., Van Rensburg H.G.;  
RT "Molecular characterization of a Galactomyces geotrichum lipase,  
RL another member of the cholinesterase/lipase family."  
RL Biochim. Biophys. Acta 1252:305-311(1995).  
CC -1- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.  
DR EMBL; X81656; CAAS7316.1; --  
DR PIR; S59958; S59958.  
DR HSP; P17573; ITHG.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR002018; CarboxylesteraseB.  
DR InterPro; IPR000379; Ser_estr.  
DR Pfam; PF00135; Coesterase; 1.  
DR PROSITE; PS00122; CARBOXYLESTERASE_B_1; 1.  
DR PROSITE; PS00941; CARBOXYLESTERASE_B_2; 1.  
KW Hydrolase.  
FT NON_TER  
SQ SEQUENCE 544 AA; 59343 MW; 688072A3E43EEB39 CRC64;  
  
Query Match 37.7%; Score 1081; DB 3; Length 544;  
Best Local Similarity 43.1%; Pred. No. 3.3e-67;  
Matches 236; Conservative 72; Mismatches 195; Indels 44; Gaps 9;  
  
QY 15 PTKLANGDITITGLNAINEAFGLIPAEPPVGNLRKDPVPYSGSLNGKFTSYGSPSCM 74  
DB 3 PTVANGNEVISGLVGKVDTFKGIIPADPPVGLRFPKFPPTGSGYQGLKANDFSSACM 62  
QY 75 QONPEGTFFE-----ENLGTALDLVMQSKVFOAVLPQSEDCLTINVRPPT 121  
DB 63 QLDPGNAISLLDKVVGKGIIPDLNLRGLPYDMA-----QGTVMSEDCLYLVNFRPAGT 116  
QY 122 KAGANLPMVLWIFGGFEGISPTIFFPPAQMTKSLVLMKGIHIVAVNRYVASWGFLAGDD 181  
DB 117 KPDAKLPVMWIIYGAFFVGSSASYPGNGYKESVEMGQPVVFSINVRTGPGYFLGGDA 176  
QY 182 IKAGSGNAGLKQORLQWVADNIAGFGDPSKVTIFGESAGSMVSLCHLIWNGDNTY 241  
DB 177 ITAEGNTNAGLHDQKGLWVSDNIANFGGDPDKVMIFGESAGAMSAHQIIVAYGGDNTY 236  
QY 242 KGKPLFRAGIMQSGAMVPSDPVDGTGNEYIYDLVFSSAGC---SASDKLACLSASD 298  
DB 237 NGKQLFHSAILQSGPLPYFSTSVGPESASVSRFAQYAGCDTSVSDNDTLACLRSKSDV 296  
QY 299 LLDATNN-----TPGFLAYSSLRLSYLPRDGKNITDDMYKLVDRDKYASVPVITG 349  
DB 297 LHSQNSYDLKDLFGLLPQFLGFG-----PRPDGNIIPDAAYELYSRGYAKVPYITG 349  
QY 350 DONDEGTIFGLSSLVNTTNAQARAYFKQSFTHASDAEIDTLMAAYPQDITQSPFDITGV 409  
DB 350 NODEGTILAPVAINATTPHVKWLYICSEASVSLDRVLSLYPGSWSEGAPFRITGIL 409  
QY 410 NALTPQFKRIASVLDGLAFIHARRYFLNHFGGKTSFLSKQLSGL--PIMGTFHNDITW 468  
DB 410 NALTPQFKRIAAITFDLLFQSPRRVLMNATKDVNRWTVLATQLHNLVFLGTGTHGSDLLF 469  
QY 469 QDYL--LGSGSVIYNNAFAFATDLPNTAGLLVNNPKYTSQSGNNLMNINAGLYTGK 527  
DB 470 QYYVDLGFSSA--YRRYFISFANHDPNVGTNLKQWDMYT---DAGKEMLIQIHMGNSMRT 525  
QY 528 DNFRTAG 534  
DB 526 DDFRIEG 532  
  
RESULT 7  
Q00892 PRELIMINARY; PRT; 544 AA.  
ID Q00892  
AC Q00892;
```


Db 350 NQDEGTAFAPVALNATTTTHVKKWLOIIFYDASEASIDRVLSLYPOTLSVSGSPFRTGIL 409
 QY 410 NALTPQFKRISAVGLDAFIHARRYFLNHPQGGTKYSFLSKQLSGL-PINGTEHANDIWM 468
 Db 410 NALTPQFKRVAALISDLMLFQSPRRVMSATKDVNRWYLSLTHLNLVPLFFGTGHENELIF 469
 QY 469 QYLLGSGSVIYNNAFIATDLPNTAGLLVNWPKYTSQSGNNIMINALGLYTGKD 528
 Db 470 QFNVTNGPANSYLRYFISFANHDPNVGTLLQWDQVT---DEGKEMLEIHMTDNVRMTD 526
 QY 529 NERTAGYDALMTN 541
 Db 527 DYRIEIGISNFETD 539
 RESULT 11
 Q8PN60 PRELIMINARY; PRT; 544 AA.
 AC Q8PN60;
 DT 01-OCT-2002 (TREMBLrel. 22, Created)
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Carboxylesterase, type B.
 GN XAC1213.
 OS Xanthomonas axonopodis (pv. citri).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 OX NCBI_TaxID=92829;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=306 / ATCC 13902 / XV 101;
 RX MEDLINE=22022145; PubMed=12024217;
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
 RA Camarotte G., Cannavan F., Cardoso J., Chambergio F., Ciapina L.P.,
 RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
 RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
 RA Setubal J.C., Kitajima J.P.;
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing
 RT host specificities";
 RL Nature 417:459-463(2002).
 DR EMBL; AB011751; AAC36085.1; --
 DR GO; GO:0005874; C:microtubule; IEA.
 DR GO; GO:0004182; F:carboxypeptidase A activity; IEA.
 DR GO; GO:0003824; F:catalytic activity; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0007018; F:microtubule-based movement; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR002453; Beta tubulin.
 DR InterPro; IPR002018; Carboxylesterase.
 DR InterPro; IPR000834; Peptidase_M14.
 DR InterPro; IPR000379; Ser.estr.
 DR Pfam; PF00135; Coesterase; 1.
 DR PROSITE; PS00122; CARBOXYLESTERASE_B_1; 1.
 DR PROSITE; PS00941; CARBOXYLESTERASE_B_2; 1.
 DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
 DR PROSITE; PS00228; TUBULIN_B_AUTOREG; 1.
 KW Complete proteome.
 SQ SEQUENCE 544 AA; 57705 MW; 55CE24A079F17BFB CRC64;
 Query Match 18.5%; Score 531; DB 16; Length 544;
 Best Local Similarity 30.2%; Pred.No.9.7e-29;
 Matches 152; Conservative 70; Mismatches 192; Indels 90; Gaps 16;

QY 25 ITGLNAINIEAFILGIPFAEPPVGNLRFKDPVPYSGSLNGKFTSYGSCMQNPEGTFEE 84
 Db 49 VRCVASAQGRAFLGVFFAAPPVGVGLAFRAFPQPPAAWTQVEDATQAGPACLPYRVG--- 104
 QY 85 NLGKTALDLVMSQKVPQAVLPQSEDCLITINVRPPGTTKAGANLPVLMWIFGGFEGISPT 144
 Db 105 -----QKHV-----SECLTLNVYAPPGPAPAHPRVAVMVWYIGGALELGSNV 146
 QY 145 IEPFAQWTKSVLMGKHIIHVAVNYRVASWGFLAGDDIKAEAGSGNAGLKDDORLGMQWVAD 204
 Db 147 DYDLSALAAR-----QDVIIVAPNYRLGVFGFYAHFGLRGEGEGVALLDQQAALRWVOR 201
 QY 205 NIAGFGGDPFKVTIFGESAGMSVLCHLIWNGDNTYKGPLFRAGIMOSGAMVPDSD--- 261
 Db 202 NIAAFGGDAHNTVFCESAGANSI-CYQLASPG-----AAGLQRAILQSGCLASDSSV 255
 QY 262 PVDCTYGNIIYDLFVSSAGCGSASDKLACLRASSDTLDDATNTTGFPLAYSSLRSLYLP 321
 Db 256 POREAESGGV--RMAQSLGCAHAPDAACLRALPADTLADAAPORRGLTGSD---AWAP 309
 QY 322 RPDGKNITDDMYKLVDRDGKVASVPVLIIGQNDGTFG-----LSSLNVTNNAQARAYFK 376
 Db 310 MSGGQALPLPPAAAIASGRHVQPVLMGTRNDEGRUFAQLLSYIGKLNLRGYEARV--- 366
 QY 377 QSFIIHASDAEI-----DTLMAAYPQDITQGSPPDTGVNLNALTQPKRISAVLGDLA 427
 Db 367 -QRMHASPAPVLRQYAAVAQAQSRWEAFADIVTDGG-----FACPTRLGRAL 412
 QY 428 FIHARRYFLNHPQGGTKYSFLSKQLSGLPIMGTFHANDIWM---QDYLLGSGSVIYNNAF 484
 Db 413 RTHAPVYAYEFDPPHAPYGLLRLPFS--PALGAFHASELVYLFQRPWVL-SGRPQFSPAQ 469
 QY 485 IAFATDL-----DPNTAG 497
 Db 470 QAFANTLDYWGAFARTGDPNGGG 493
 RESULT 12
 Q90ZK8 PRELIMINARY; PRT; 603 AA.
 AC Q90ZK8;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Butyrylcholinesterase precursor (EC 3.1.1.8).
 GN BCHE
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]_TaxID=9031;
 RP SEQUENCE FROM N.A.
 RA Geisler K., Chatonnet A., Layer P.G.;
 RT "Chicken Butyrylcholinesterase";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.
 DR EMBL; AJ306928; CAC37792.1; --
 DR GO; GO:0004104; F:cholinesterase activity; IEA.
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR InterPro; IPR002018; Carboxylesterase.
 DR InterPro; IPR000997; Cholinesterase.
 DR InterPro; IPR000379; Ser.estr.
 DR Pfam; PF00135; Coesterase; 1.
 DR PRINTS; PR00878; CHOLINESTRASE.
 DR PROSITE; PS00122; CARBOXYLESTERASE_B_1; 1.
 DR PROSITE; PS00941; CARBOXYLESTERASE_B_2; 1.
 KW Hydrolase; Signal.
 FT SIGNAL 1 29 POTENTIAL.
 FT CHAIN 30 603 BUTYRYLCHOLINESTERASE.
 SQ SEQUENCE 603 AA; 68480 MW; A350FDDF68574ADF CRC64;
 Query Match 17.3%; Score 495; DB 13; Length 603;

Best Local Similarity 30.5%; Pred. No. 3.7e-26;
Matches 163; Conservative 70; Mismatches 199; Indels 102; Gaps 22;

QY 22 GDTITGLNAINAFGIPFAEPVGNLRPKDPVPYSGSLNGKQFTSYGSCWQONDEGT 81
DB 50 GGTVT-----AFLGIPYKGPGRGRFQKPEPFKWSGIWKATKHANSCYQ----- 95

QY 82 FEENLGTALDLVMQSKVFOAVLPQSEDCLTINVRP-PGTKAGANLPVMLWIFGGGFEI 140
DB 96 ----LIDTTPGPGTEWMPKNTLSEDCLYLNWIIPSPKP-----NATVWVWIIYGGSFET 148

QY 141 GSPTIFPPAQWTKSVLMGK-----HIIHVAVNRYVASWGFAGDDDKAEGSGNAGLKD 194
DB 149 GSTSL-----PVYDGRFLARVERVIVVMNRYTGALGFLALPGNK-EVPGNAGLFD 198

QY 195 QRLGMQWVADNIAGFGDPSKVTIFGSGAGSMVLCILWINDGNTYKGPLFRAGIMQS 254
DB 199 QRLALQWQENIASFGGNPKSVTIFGSGAGSASVSYHIL-----SPKSHPLFTRAIMQS 252

QY 255 G-AMVPSDPVDGTGNEIYDLFVSSAGCGSA--SDKLACLRSSASDITLLDATNTPGFLA 311
DB 253 GSANAPWAAITASARRRTVALAKQLKCPSTDETELILCLQDKDPKDILE--NEVYVVKY 310

QY 312 YSSRLSYLPRPDGKNITDDMYKLVRDGYKASVPVVIIGDQDEGTIF-----GLS--SL 363
DB 311 FSLHIIYFCPTVGDGFLADMPEALIKNGIPKQTVLVGNKDEGTSFLVYGVPGFSKDS 370

QY 364 NVTNAQARAYFKOSFTHASDAEIDTLMAAY-----PQDITQSGSPFDGVLNALT-P- 414
DB 371 SLINKTOPEVALTISFPQVSKLAIESIFQYTDWENQKPEHYRDAMDVDVIGDHIICPA 430

QY 415 -QFKRISAVLGLAFIHARRYFLNHFQGTGYKVSFLSKOLSLPTMGTFHANDIVWDYLL 473
DB 431 VEFKTAIEVGNVFF-----YFFEI-----RSSKLPWPEWGVHGEYI---BEVF 474

QY 474 G---SGSVIYNA-----FIAPATLDLPNTAGLVN---WPXYTSSQ 510
DB 475 GLPLERRVNYTKAEILSRSLRYWASFAKTGNPN--GTLINGTRWPVFTSTEQ 526

RESULT 13
Q9XVA9
ID O9XVA9 PRELIMINARY; PRT; 656 AA.
AC O9XVA9
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Acetylcholinesterase (Fragment).
GN ACHE.
OS Meloidogyne javanica (Root-knot nematode).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
OC Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.
OX NCBI_TaxID=6303;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99270313; PubMed=10340488;
RA Plette C., Arthaud L., Abad P., Rosso M.N.;
RT "Molecular cloning of an acetylcholinesterase gene from the plant
parasitic nematodes, Meloidogyne incognita and Meloidogyne javanica.";
RL Mol. Biochem. Parasitol. 99:247-256(1999).
CC -!- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.
DR EMBL; AF080184; RAD25921.1; --
DR HSP; P21836; LMAA.
DR GO; GO:0004104; F:cholinesterase activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR002018; Carboxylesterase.
DR InterPro; IPR000997; Cholinesterase.
DR InterPro; IPR001064; Crystallin.
DR InterPro; IPR000379; Sex esters.
DR Pfam; PF00135; Coesterase; 1.
DR PRINTS; PR00878; CHOLNESTRASE.
DR PROSITE; PS00122; CARBOXYLESTERASE_B_1; 1.
DR PROSITE; PS00941; CARBOXYLESTERASE_B_2; 1.
DR PROSITE; PS00941; CARBOXYLESTERASE_B_2; 1.

DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
KW Hydrolase. 656 656
FT NON TER 656 AA; AF90A221AEBFF00C CRC64;
SQ SEQUENCE 656 AA; 76061 MW; 76061 MW; AF90A221AEBFF00C CRC64;

Query Match 16.9%; Score 484; DB 5; Length 656;
Best Local Similarity 27.6%; Pred. No. 2.5e-25;
Matches 156; Conservative 90; Mismatches 205; Indels 114; Gaps 23;

QY 19 LANGDITGLNATINEA-----FLGIPFAEPVGNLRPKDPVPYSGSLNGKQFTSYG 71
DB 49 LSDGSPIFG--SFQAATGKYISQFLGVPAEPPIGKLRFRPIPKRPWEQWNAITFRD 106
QY 72 SCMQQNDPEGTFEENLGTALDLVMQSKVFOAVLPQSEDCLTINVRPPTGKAGANLPVL 131
DB 107 SCV-QSPDTYFGFYGAT-----MNSNTPCEDCLYNTIYVPEIDREKRLPVLF 156

QY 132 WIFGGGFEIGSPITFPFAQWTKSVLMG-KHIIHVAVNRYVASWGF-AGDDIKAECSGN 189
DB 157 WIYGGGFWSGTASL-----DVYDGKIFAGEENVIIIVYRVTVFGFYLIG---REAPGN 209

QY 190 AGLKDQRLGMQWVADNIAGFGDPSKVTIFGSGAGSMVLCILWINDGNTYKGPLFRA 249
DB 210 MGLWDQLLALKWYKNIQVFGDPSLITLFGESAGASVSMHML-----SPLSQPVFTR 263

QY 250 GIMQSGAMVPSDPVDGTGYNE-----IYDLFVSSAGCGSAS-----DKLAC 290
DB 264 SILQSGAATAPWAVE-----NKQVALHRAVILYEYV--KCGNGNMHSLAPDOWNMDEVLC 317

QY 291 LRSASDITLDATNTPGFLAYSSLR-----LSYLPRDQKNITDDMYKLVRDGVKVASVPV 346
DB 318 LHAASADKLKRS-----EWSVMEFADFPPWVFVDGEFLVENIETSLKRGNFKKTLQ 369

QY 347 IIGDNDGTTIFGLSSLNVTNAQARAYFKOSFTHASDAEIDTLMAAYPQDITQSGPDT 406
DB 370 LAGSNFDEAIYIVYQLS-DVFPFAEPFEKKDFIKSRDVMIKNVSSLLPRQILKSSLAQ 428

QY 407 GVLNALTPO-----FKRISAVLGLDLP-IHARRYFLNHFQGT-----KYSFLSKQ 451
DB 429 SILNYEPEGLPIESKNWVDSLDMGLDFLTCNVNEFALAHSEHGADTYTYMFESHASQ 488

QY 452 LSGLPIMGTFFHANDIVWDYLLGSSSVIYNA-----FIAPATLDLP 493
DB 489 QTPWPEWGVHGEYI---NFIFGP---YNRKQFKYKEQEQLSSRFMRFWANPARTSDP 542

QY 494 N---TAGLLVNMWPKYTSSSQSGNNL 515
DB 543 MHPNPNYSISDPYNSKTMVEYNL 567

RESULT 14
O96529
ID O96529 PRELIMINARY; PRT; 656 AA.
AC O96529;
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Acetylcholinesterase.
GN ACHE.
OS Meloidogyne incognita (southern root-knot nematode).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
OC Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.
OX NCBI_TaxID=6306;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99270313; PubMed=10340488;
RA Plette C., Arthaud L., Abad P., Rosso M.N.;
RT "Molecular cloning of an acetylcholinesterase gene from the plant
parasitic nematodes, Meloidogyne incognita and Meloidogyne javanica.";
RL Mol. Biochem. Parasitol. 99:247-256(1999).
CC -!- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.
DR EMBL; AF075718; AAD02835.1; --
DR HSP; P21836; LMAA.

Blank

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 2004, 14:51:20 ; Search time 4047 seconds
(without alignments)
16407.593 Million cell updates/sec

Title: US-09-943-857-3

Perfect score: 1532

Sequence: 1 gccaccgcgaagctcgcca.....tgaccacccgtctttgtg 1532

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.in.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pln.*

35: em.htg.rod.*

36: em.htg.man.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1532	100.0	1532	6	AX709921	AX709921 Sequence
2	1271.6	83.0	1855	8	CCLIP3	X66006 C.cylindrac
3	1191.6	77.8	1647	8	AY464467	AY464467 Candida c
4	1080.6	70.5	1541	6	AX709925	AX709925 Sequence
5	1025	66.9	1657	8	CCLIPASE	X16712 C. cylindra
6	1022	66.7	1733	6	A48377	A48377 Sequence 6
7	1022	66.7	1733	6	CCLIP1	X64703 C.cylindrac
8	1007.6	65.8	1511	6	AX709927	AX709927 Sequence
9	1002.4	65.4	1647	6	A81171	A81171 Sequence 1
10	998.4	65.2	1548	6	AX709923	AX709923 Sequence
11	982.4	64.1	1812	8	CCLIP5	X66008 C.cylindrac
12	937.2	61.2	1785	8	CCLIP4	X66007 C.cylindrac
13	867.6	56.6	2043	8	CCLIP2	X64704 C.cylindrac
14	830.4	54.2	1469	6	AX709919	AX709919 Sequence
15	831.8	42.5	1688	12	AF044078	AF044078 Synthetic
16	590.2	38.5	1605	6	A81172	A81172 Sequence 2
17	173.2	11.3	1635	8	GU02541	U02541 Geotrichum
18	173.2	11.3	1635	8	GU02625	U02625 Geotrichum
19	173	11.3	2045	6	A45569	A45569 Sequence 7
20	173	11.3	2045	6	A47607	A47607 Sequence 7
21	173	11.3	2045	6	AR019333	AR019333 Sequence
22	172.8	11.3	1635	8	GU02623	U02623 Geotrichum
23	171.2	11.2	1674	6	E02497	E02497 cDNA encodi
24	171.2	11.2	1767	8	GCALIP2	D0697 Geotrichum
25	171.2	11.2	1783	8	AB000260	AB000260 Geotrichu
26	164.2	10.7	1635	8	GCTAGL	X81556 G.candidum
27	164.2	10.7	1635	8	GU02525	U02525 Geotrichum
28	164.2	10.7	1828	6	A74255	A74255 Sequence 11
29	161.8	10.6	1635	6	A48376	A48376 Sequence 5
30	161.8	10.6	1635	8	GU02387	U02387 Geotrichum
31	159.4	10.4	1635	8	GU02622	U02622 Geotrichum
32	157.8	10.3	1692	6	E02678	E02678 cDNA encodi
33	156.2	10.2	1635	8	GU02524	U02524 Geotrichum
34	128.4	8.4	1934	8	AK107259	AK107259 Oryza sat
35	125.2	8.2	1692	8	GGECLR	X78032 G.geotrichu
36	120	7.8	4295	8	AY250996	AY250996 Aspergill
37	88.2	5.8	1716	6	AX268011	AX268011 Sequence
38	88.2	5.8	1738	6	AR034144	AR034144 Sequence
39	88.2	5.8	1738	6	BD005938	BD005938 Cephalosp
40	88.2	5.8	1738	8	AF025410	AF025410 Rhodospo
41	87	5.7	2000	8	ANGIAA5	X56442 A.niger gla
42	70.8	4.6	13546	1	AE011751	AE011751 Xanthomon
43	70.2	4.6	6028	3	AF177382	AF177382 Culex tri
44	65.2	4.3	1520	3	CQSERES3	Z32696 C.quinquefa
45	65.2	4.3	1836	3	CQSERES1	Z32694 C.quinquefa

ALIGNMENTS

RESULT 1	AX709921	Sequence 3 from Patent EP1288294.	1532 bp	DNA	linear	PAT 10-APR-2003
LOCUS	AX709921	Sequence 3 from Patent EP1288294.				
DEFINITION	AX709921	Sequence 3 from Patent EP1288294.				
ACCESSION	AX709921	Sequence 3 from Patent EP1288294.				
VERSION	AX709921.1	GI:29786302				
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						

Academia Sinica (TW)		Location/Qualifiers	
source		1..1532	
FEATURES		/organism="Candida rugosa"	
		/mol_type="unassigned DNA"	
		/db_xref="taxon:5481"	
ORIGIN			
Query Match 100.0%; Score 1532; DB 6; Length 1532;			
Best Local Similarity 100.0%; Pred. No. 6.1e-243;			
Matches 1532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	GCCACCGCCCAAGCTCGCAACGCGGACACCATCACCGGTCTCAACGCGCATCATCAACGA	60
DB	1		
QY	61	GCGCTTCTCGGCAATCCCTTTGCGAGCCGCGCGTGGGCAACCTTCGCTTCAAGGACCC	120
DB	61		
QY	121	TGTGCGTACTTGTGCTCAAGCGGCGAGAGTACTTACGGCGCGTGCATGCAGCAG	180
DB	121		
QY	181	AACCCGAGGGCAGCTTTGAAGAGAACTTGGCAAGCGGCACTCGACTTGGTGATGCAG	240
DB	181		
QY	241	TCCAAAGTGTTCCAGCGCGTGTCTCCCGAGAGTGGACTGCTCACCATCAACGTGGTG	300
DB	241		
QY	301	CGGCGCGGGGACCAAGCGGGCGCAACTCCCGGTCATGCTCTGGATCTTTGGCGGT	360
DB	301		
QY	361	GGGTTTGAGATCGGAGCGCCCACTCTCCCGCGGCGCAACTCCCGGTCATGCTCTGGATCTTTGGCGGT	420
DB	361		
QY	421	CTCATGGCAAGCCCATCATCCAGTGGCGGTCAACTACGGTGTTCCTGCTGGGGGTCT	480
DB	421		
QY	481	TGGCTGTGTACATCAAGCGCGGAGCGGCGGCAAGCGCGGCTTGAAGGACCAAGCTT	540
DB	481		
QY	541	TGGGATGCAATGAGTGGGAGCAACATTTGCCGGGTTCCGGCGGACCCGAGCAAGGTGA	600
DB	541		
QY	601	CATCTTGGGAGGCGGCGAGCATGTCCTGTGTGGCCACCTCATCTGGAACGACGGCGA	660
DB	601		
QY	661	CAACGCTACAGGCGCAAGCGGTTCGCGCGGCGCATCATGCGAGGAGCCATGCTGCC	720
DB	661		
QY	721	GGACCCGTTGAGCGGCACTAGCGCAACGAGATCTACGACCTTTTGTCTCGAGTGTGG	780
DB	721		
QY	781	CTGTGGAGCGCCAGGCAACAGCTGCGTGTGGCAGTGGGAGGACACCTTGTCTCA	840
DB	781		
QY	841	TGCCACCAACACTCTTGGGTCTTGGCGTACTCTCTGTTGGGTGTACTCTCCGGCC	900
DB	841		
QY	901	CGACGCGAAGAAATCACCGATGACATGTACAAGTTGGTGGCGGCGCAAGTATGCAAG	960
DB	901		
QY	961	CGTTCCGCTGATCATTTGGCGACGAGGAGGCGACCATCTTTGGCTCTTGAACGTG	1020
DB	961		
QY	1021	ACCAAGATGCTCAGGCGCGTGTACTTCAAGCAGTTTCAATCCAGCGCAGCGAGCGGAG	1080
DB	1021		
QY	1081	ATCGACACCTTGTATGGGCGGTACCCCGAGGAGTCAACCGAGGTCCTTTCGACACGGTT	1140
DB	1081		
QY	1141	CAACGCTCACCCCGAGTTCAAGAGATCGCGGTCTCGGCGACCTTCATTCATCCACG	1200
DB	1141		
QY	1201	CCCGCGCTACTTCTCAACCACTTCCAGGGCGGCAACCAAGTACTCTTCCTCAAGCAGC	1260
DB	1201		
QY	1261	TCGGGTTGCCAATCATGAGGCGACCTTCCATGCGCAACGACATTTGTGGCAGGACTACTTGT	1320
DB	1261		
QY	1321	TGGGAAGCGGAGCGGTCTACACCAACGCGTTTATCGGTTCCGCCACCGACTTTGGACC	1380
DB	1321		
QY	1381	CCAACACCGCGGGTGTGTGGTGAACCTGCCCAAGTACACGAGCAGCGGCAACAAC	1440
DB	1381		
QY	1441	TGATGATGATCAACGCTTGGGCTTGTATACCGGCAAGGACACTTCGCGACCGCTGGCT	1500
DB	1441		
QY	1501	ACGACGCTTGTATGACCAACCGCTTCTTTGTG	1532
DB	1501		
RESULT 2			
CCLIP3 1855 bp DNA linear PLN 30-JUN-1993			
LOCUS C.cylindracea LIP3 gene.			
DEFINITION X66006 S55937			
ACCESSION X66006.1 GI:296933			
VERSION lip3 gene; lipase.			
KEYWORDS Candida cylindracea			
SOURCE Candida cylindracea			
ORGANISM			
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
Saccharomycetales; mitosporic Saccharomycetales; Candida.			
REFERENCE 1 (bases 1 to 1855)			
AUTHORS Lotti,M., Grandori,R., Fusetti,F., Longhi,S., Brocca,S.,			
Tramontano,A. and Alberghina,L.			
Cloning and analysis of Candida cylindracea lipase sequences			
Gene 124 (1), 45-55 (1993)			
JOURNAL 93178975			
MEDLINE 840480			
PUBMED 2 (bases 1 to 1855)			
REFERENCE Longhi,S.			
AUTHORS Direct Submission			
JOURNAL Submitted (06-MAY-1992) S. Longhi, Universita' degli Studi di			
Milano, Via Celoria 26, 20133 Milano, ITALY			
COMMENT See X66006-8, X16712, X64703 & X64704			
X66006, X66007 and X66008 are related sequences in the description			
of C. cylindracea Lip1 and Lip2 sequences (x64703, x64704).			
FEATURES			
Location/Qualifiers			
1..1855			
/organism="Candida cylindracea"			
source			

```

/mol_type="genomic DNA"
/strain="ATCC 14830"
/db_xref="taxon:44322"
/clone_lib="SACI genomic"
84. .1733
/gene="LIP3"
/gene="LIP3"
84. .1733
/gene="LIP3"
/EC_number="3.1.1.3"
/codon_start=1
/product="triacylglycerol lipase"
/protein_id="CAA45805.1"
/db_xref="GI:1325389"
/db_xref="GOA:P32947"
/db_xref="SWISS-PROT:P32947"
/translation="MKALALILIASVAAPTAKLANGDTITGLNAIINEAFLGIPFA
EPFNNLRFPKDPVPYSGSLNGQKFTLYGPLCQONPEGTPEENLGKTLALVMQSKVFI
QAVLPQSDGCTINVRPPEGTKAGANLPVLMWIFGGFEGISPTI FPPAQMTVKSLM
CKPIITHAVNVAVRVAASPGFLAGDDIKAEGSNAGIKDORLGMQWADNIAAGFGDPKVI
TIFGLAGSMVYLCHLIWNDSGNTYKGPFLFRAGIMQLGANPLDPVDGTYGNIYIDL
FVSSAGGSASGLKALNSALSDTLDDNNTPGFLAYSSIRLRLYLRPPQKNDITDDM
YKLVNRGKYASVPVLIIGNDSGTYGLLLNNVTNQAARAYFQKLFIFHSDAEIDITL
MAAYPQDITQGLPFTGTFNATITPQFKRILAVLGDLPFIHARRYFLNHFQGTGYISFL
LKQLGLPMIGMTFHANDIYWDQYLLGSGSVIYNNAFATAFATDLDLPNTAGLLVNWPKYTS
SSLQLGNLMMINALGLYTGKDNFRTAGYDALMTNPLLIFFV"
84. 218
/gene="LIP3"
219. .1730
/gene="LIP3"
/product="triacylglycerol lipase"
/EC_number="3.1.1.3"
sig_peptide
mat_peptide

```

ORIGIN

Query Match	83.0%;	Score 1271.6;	DB 8;	Length 1855;
Best Local Similarity	94.9%;	Pred. No. 5.1e-200;		
Matches 1505;	Conservative 0;	Mismatches 19;	Indels 62;	Gaps 16;
Qy	2	CCACGCCCAAGCTCGCCAAAGCGGACACCATCATCCGCTCTCAACGCCATCATCAACGAG	61	
Db	132	CCACGCCCAAGCTCGCCAAAGCGGACACCATCATCCGCTCTCAACGCCATCATCAACGAG	191	
Qy	62	GCCTTCCTCGGCATTCCCTTTGCCGAGCGCGCGTGGGCAACCTCCGCTTCAAGGACCCCT	121	
Db	192	GCCTTCCTCGGCATTCCCTTTGCCGAGCGCGCGTGGGCAACCTCCGCTTCAAGGACCCCT	251	
Qy	122	GTGCGGTACTCTGGCTCGCTCAACGGCGGCAAGTTACTTT-----ACGGCCCGGTGATG	174	
Db	252	GTGCGGTACTCTGGCTCGCTCAACGGCGGCAAGTTACTCTGTGACGGCCGCGCTGTGCATG	311	
Qy	175	CAGCAGAAACCCGAGGGCACGTTTGAAGAGAACTCTGGCAAGCGGCACCTCGACTTCGGTG	234	
Db	312	CAGCAGAAACCCGAGGGCACGTTTGAAGAGAACTCTGGCAAGCGGCACCTCGACTTCGGTG	371	
Qy	235	ATGCAGTCCAAGGTGTTCCAGCGCGTGTCTTCCAGAGAGTGAAGGACTGCCTCACCATCAAC	294	
Db	372	ATGCAGTCCAAGGTGTTCCAGCGCGTGTCTTCCAGAGAGTGAAGGACTGCCTCACCATCAAC	431	
Qy	295	GTGGTGGCGCGCCGGGACAACGAGCGGGCGCCAACTCCCGGTGATCTCTGATCTTT	354	
Db	432	GTGGTGGCGCGCCGGGACAACGAGCGGGCGCCAACTCCCGGTGATCTCTGATCTTT	491	
Qy	355	GGCGGTGGTTTGAGATCGGCAGCCCAACATCTTCCCTCCCGCCAGATGGTCACCAAG	414	
Db	492	GGCGGTGGTTTGAGATCGGCAGCCCAACATCTTCCCTCCCGCCAGATGGTCACCAAG	551	
Qy	415	AGTGTGCTCATGGGCAAG-CCATCATCAACGTGGCGGTCAACTACCGTGTTCCTCGTG	473	
Db	552	AGTGTGCTCATGGGCAAGCCCATCATCAACGTGGCGGTCAACTACCGTGTTCCTCGTG	611	
Qy	474	GGGTTCCTTGGCTGTGTATGATCAATCAAGGCGAGGGCAGCGGGAAACGCCGGCTTGAAGAC	533	
Db	612	GGGTTCCTTGGCTGTGTATGATCAATCAAGGCGAGGGCAGCGGGAAACGCCGGCTTGAAGAC	671	

```
RESULT 3
AY464467      1647 bp      mRNA      linear      PLN 03-DEC-2003
LOCUS
DEFINITION   Candida cylindracea lipase mRNA, partial cds.
ACCESSION   AY464467
VERSION      AY464467.1  GI:38565535
KEYWORDS
SOURCE
ORGANISM     Candida cylindracea
              Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
              Saccharomycetales; Mitosporic Saccharomycetales; Candida.
REFERENCE
AUTHORS      Wang,J.-Z., Wei,D.-Z. and Zhang,L.-Y.
TITLE        Cloning and sequence analysis of a lipase gene from Candida rugosa
JOURNAL
JOURNAL
AUTHORS
JOURNAL
FEATURES
source       1..1647
             /organism="Candida cylindracea"
             /mol_type="mRNA"
             /db_xref="taxon:44322"
             1..-1647
             /EC_number="3.1.1.3"
             /codon_start=1
             /product="lipase"
             /protein_id="AAR24090.1"
             /db_xref="GI:38565535"
             /translation="MELALLLIASVAAPAKLANGDTITGLNAINIENAFILPFA
             EPPVGNLPDPVPVPSGLNGKFTLYCPLOMCPQNPSTPEENIGKTLADLMVMSKVF
             QAVLPQSEDLITINVRPPTKAGANLPVLMWIFGGPEIGSPILFPPAQWTKSVLM
             KPLIHVAVNTRVSPGFLAGDDIKABSGNAGLUKQRLGMQWADNIAGPQDPFVK
             TIFELAGSLSVLCHLIWNGDNTYKGLPFRAGIMQIGAMVPLDPVDYVGNIEYDL
             FVSSAGCGSADKLALRSALSDLLDNTNTPGLFAYSLRLLYLPRPDGKNITDM
             YKLVRDGKASVPIIGQNDQGTFFGLLIANVTDAEARQYFKELFVHTSDABLDTL
             MTAYPDQITQSPDGTGLNALTPQKRLILAVLGLDAFIHARRVFLNHYGTGVYSFL
             LKQLLGLAVLTGFHNDIVFDYLYVSGSVIYNNAFTAFVTDLPNTAGLLVNWFEYT
             SSULQGNLMMINALGLYTGKDNFRTAGIDALMANPPLFFV"
ORIGIN
Query Match      77.8%; Score 1191.6; DB 8; Length 1647;
Best Local Similarity 91.4%; Pred. No. 8.1e-187;
Matches 1448; Conservative 0; Mismatches 74; Indels 62; Gaps 15;
Qy 2 CCCACGCCAAGCTCGGCAACGGCGACACCATCACCGGTCTCAACGCCATCATCAACGAG 61
Dy 49 CCCACGCCAAGCTCGGCAACGGCGACACCATCACCGGTCTCAACGCCATCATCAACGAG 108
Qy 62 GCGTTCCTCGGATTCCTTTCCGAGCGCGGCTGGGCAACCTCCGCTTCAAGGACCCCT 121
Dy 109 GCGTTCCTCGGATTCCTTTCCGAGCGCGGCTGGGCAACCTCCGCTTCAAGGACCCCT 168
Qy 122 GTGCGGTACTCTGGCTCGCTCAACGGCGCAGAGTACTT-----ACGGCCCGGTGATG 174
Dy 169 GTGCGGTACTCTGGCTCGCTCAACGGCGCAGAGTACTTCTGTAGCGGCCGCTGTGTCATG 228
Qy 175 CAGCAGAAACCCCGAGGGCACTTTGAAGAGAACCTTTGGCAGAGACGGCACTCGACTGTGGTG 234
Dy 229 CAGCAGAAACCCCGAGGGCACTTTGAAGAGAACCTTTGGCAGAGACGGCACTCGACTGTGGTG 288
Qy 235 ATGCAGTCCAGGTTTCCAGCGGTGCTTCCAGAGTGAGGAGTGCCTCACCATCAAC 294
Dy 289 ATGCAGTCCAGGTTTCCAGCGGTGCTTCCAGAGTGAGGAGTGCCTCACCATCAAC 348
Qy 295 GTGGTGGCGCGCGCGGACCAAGCGCGCGCAACCTCCCGGTGATGCTCTGGATCTTT 354
Dy 349 GTGGTGGCGCGCGCGGACCAAGCGCGCGCAACCTCCCGGTGATGCTCTGGATCTTT 408
Qy 355 GGCCTGGGTTTGAATCGGCAAGCCCAACATCTTCCCTCCCGCCAGATGTCACCAAG 414
Dy 1489 ACCGCGGGGTTGCTGGTGAATCGGCAAGCCCAACATCTTCCCTCCCGCCAGATGTCACCAAG 1548
```

```
Db 409 GGCCTGGGTTTGAGATCGGCAAGCCCAACCATCTTCCCTCCCGCCAGATGGTCAACCAAG 468
Qy 415 AGTGTGCTCATGGCAAG--CATCATCAAGTGGCCCGTCAACTACCGTGTGCTCGTGG 473
Db 469 AGTGTGCTCATGGCAAGCCCATCATCAAGTGGCCCGTCAACTACCGTGTGCTCGTGG 528
Qy 474 GGGTTCTTGGCTGGTGTATGATCAAGGCCGAGGGCAGCGGGAACGCGCGGTGGAAGAC 533
Db 529 GGGTTCTTGGCTGGTGTATGATCAAGGCCGAGGGCAGCGGGAACGCGCGGTGGAAGAC 588
Qy 534 CAGGTTTGGCATGAGTGGTGGCAGACAAATTCGCGGGTTCGGGGCGGACCCGAGC 593
Db 589 CAGGTTTGGCATGAGTGGTGGCAGACAAATTCGCGGGTTCGGGGCGGACCCGAGC 648
Qy 594 AAGGTGAC-ATCTTTGGCGA---GGCGGGCAGCATGTCCGTTGTGTGCACTCATCTGG 649
Db 649 AAGGTGACGATCTTTGGCGAGCTGGCGGGCAGCATGTCCGTTGTGTGCACTCATCTGG 708
Qy 650 AACGACGGCGACAAACACGTACAAGGGCAAGCGGTGTTCGCGCGGGGATCATGCA---G 706
Db 709 AACGACGGCGACAAACACGTACAAGGGCAAGCGGTGTTCGCGCGGGGATCATGCACTG 768
Qy 707 GGAGCCATGTGGC---GGACCCCGTGGAGGGCAGCTACGGCAACGAGATCTAGACCTC 763
Db 769 GGAGCCATGTGGCCTGGACCCCGTGGAGCGGACGTCAGCAACGAGATCTAGACCTC 828
Qy 764 TTTGCTCGAGTGTGCTGTGGCAGCGCCAGCAAGCTCGCGTGTGTGCGCAGTGC- 822
Db 829 TTTGCTCGAGTGTGCTGTGGCAGCGCCAGCAAGCTCGCGTGTGTGCGCAGTGC 888
Qy 823 --GAGCGACACCTTGTCTGATGTCACCAACAACTCTCTGGGTTCCTGGCGTACTCTCTG 880
Db 889 CTGAGCGACACCTTGTCTGATGTCACCAACAACTCTCTGGGTTCCTGGCGTACTCTCTG 948
Qy 881 TTGCGGTGTACT-----CCCGCCCGAGCGGCAACATCAACGATGATGATGACAG 934
Db 949 TTGCGGTGTGTACTTCCCGCCCGAGCGGCAAGCAATCAACGATGATGATGACAG 1008
Qy 935 TTGTCGCGCAGCGCAAGTATGCAAGGTTCCTCGTATCATTTGGCGCACCAAGACGAG 994
Db 1009 TTGTCGCGCAGCGCAAGTATGCAAGGTTCCTCGTATCATTTGGCGCACCAAGACGAG 1068
Qy 995 GGCACCATCTTTG-----GCTTTGAAAGTGAACCAAGATGCTCAGGCCCGTCTTAC 1047
Db 1069 GGCACCTTCTTCGAGACTCTGCTGTGAAAGTGAACCAAGATGCTCAGGCCCGTCTTAC 1128
Qy 1048 TTCAAGCA--GTTTCATCCAGCGCAGCGGAGATCGACACTTGATCGCGCGCTAC 1104
Db 1129 TTTAAGAGTGTGTTGTCCACCAAGCGATGCGGAGCTCGACAGTGTATGACGCGCTAC 1188
Qy 1105 CCCCAGGACATCACCCAGGG---TCCGTTTCGACACGGTT-----CAACGCTCAACCCG 1154
Db 1189 CCCCAGGACATCACCCAGGGCTCGCGTTTCGACACGGGTATTTCTCAAGCCCTCAACCCG 1248
Qy 1155 CAGTTCAGAGAAATC---GGGTGCTCGCGCACTTGCATTCATCCAGCGCGCGCTAC 1211
Db 1249 CAGTTCAGAGAAATCCTTGGCGGTGCTCGCGCACTTGCCTTCATCCACCGCGCGCTAC 1308
Qy 1212 TTCTCTACACACTTCCAGGGCGGACCAAGTACTCGTTCCTC-----AAGCAGTCTGG 1265
Db 1309 TTCTCTACACACTACACCGGGCGGACCAAGTACTCGTTCCTCCTTGAAGCAGCTCTGG 1368
Qy 1266 TTGCCAATCATGGGCACCTTCCATGCCAACGACATTTGTGTGGCAGGACTACTTGTGGGA 1325
Db 1369 TTGGGGTGTCTGGAAAGTTCACGCAACGACATGTTCTTCCAGGACTACTTGTGGGCG 1428
Qy 1326 AGCGCAGCGTCACTCAACAAACGCGTTTATCGGTTTCGCGCACCGACTTGGACCCCAAC 1385
Db 1429 AGCGCAGCGTCACTCAACAAACGCGTTTATCGGTTTGTCAAGGACTTGGACCCCAAC 1488
Qy 1386 ACCGCGGGTGTGTTGTTGAATGTCGCCCAAGTACACCAAGCAGC-----CAGGGCAACAC 1439
Db 1489 ACCGCGGGTGTGTTGTTGAATGTCGCCCAAGTACACCAAGCAGCCTGACGCTGGGCAACAC 1548
```


QY 1440 TTGATGATGATCAACGCTTGGGCTTGTACACGGCAAGACAACTTCCGACCGCTGGC 1499
 Db 1549 TTGATGATGATCAACGCTTGGGCTTGTACACGGCAAGACAACTTCCGACCGCGGC 1608
 QY 1500 TACGACGCGTTGATGACCAACCGC 1523
 Db 1609 TACGACGCGTTGATGCTAACCGC 1632

RESULT 4
 AX709925
 LOCUS
 DEFINITION
 Sequence 7 from Patent EP1288294.
 ACCESSION
 AX709925
 VERSION
 AX709925.1 GI:29786304

KEYWORDS
 Candida rugosa
 Candida rugosa
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE
 1
 Tang, S.J., Lee, G.C. and Shaw, J.F.
 Recombinant candida rugosa lipases
 Patent: EP 1288294-A 7 05-MAR-2003;
 Academia Sinica (TW)

FEATURES
 Location/Qualifiers
 1. .1541
 /organism="Candida rugosa"
 /mol_type="unassigned DNA"
 /db_xref="taxon:5481"

ORIGIN
 Query Match 70.5%; Score 1080.6; DB 6; Length 1541;
 Best Local Similarity 84.9%; Pred. No. 1.6e-168;
 Matches 1311; Conservative 0; Mismatches 209; Indels 25; Gaps 8;

QY 3 CCACCCCAAGCTCGCCAAAGCGGACACCATCACCGGTCTCAACGCCATCATCAACGAGG 62
 Db 2 CCACGCCAGCTCGCCAAAGCGGACACCATCACCGGTCTCAACGCCATCATCAACGAGG 61
 QY 63 CGTTCCTCGGCATCTCGCTCGCTCAACGGCCAGAGTT----ACTTACGGCCGCTGATCGAG 122
 Db 62 CGTTCCTCGGCATCTCGCTCGCTCAACGGCCAGAGTT----ACTTACGGCCGCTGATCGAG 121
 QY 123 TGGCGTACTCTGCTCGCTCAACGGCCAGAGTT----ACTTACGGCCGCTGATCGAG 178
 Db 122 TGGCGTACTCTGCTCGCTCAACGGCTCAATCTTACCGGTACGGTCCGCTGATCGAGC 181
 QY 179 AGAACCCCGAGGACGCTTTGAAGAGAACCTTGGCAAGACGGCACTCGACTTGGTATGC 238
 Db 182 AGAACCCCGAGGACGCTTACGAGAGAACCTTCCCAAGGTGGCGTTGACTTGGTATGC 241
 QY 239 AGTCCAAAGTGTCCAGGCGGTGCTTCCAGAGTGAAGACTGCTCCATCAACGAGTGG 298
 Db 242 AGTCCAAAGTGTTCAGGCTGTCTTCCCAAGAGGAGTGTCTCCCAAGAGGAGTGCCTC 301
 QY 299 TGGCGCGCGCGGACCAAGGCGGCGCCAACTTCCCGGTCACTGCTGATCTTTGGCG 358
 Db 302 TGGCGCGCGGCGACCAAGGCGGCGCCAACTTCCCGGTCACTGCTGATCTTTGGCG 361
 QY 359 GTGGGTTTGAGATCGGACGCCCAACCATCTTCCCTCCCGCCAGATGGTCAACAGAGTG 418
 Db 362 GTGGGTTTGAGATCGGACGCCCAACCATCTTCCCTCCCGCTCAGATGGTCTCCAGAGTG 421
 QY 419 TGTCTATGGCAAGCATCATCATCGTGGCGGTCAACTACCGGTGTGCTCGTGGGGTT 478
 Db 422 TGTCTATGGCAAGCATCATCATCGTGGCGGTCAACTACCGGTGTGCTCGTGGGGTT 481
 QY 479 CTTGGCTGTGATGATCAAGGCGGCGGAGCGCGGCTTGAAGACACGCG 538
 Db 482 CTTGGCGGTGCGGATCAAGGCGGAGGCGAGCTCAATGCGCGGCTCAAGGACGAGCG 541

RESULT 5

CCLIPASE

LOCUS

DEFINITION

C. cylindracea mRNA for lipase I, partial cds.

QY 539 TTTGGGCATGACGATGGGTGGCAGACAACTTGCCTGGGTTCCGGCGGACCGCAAGGT 598
 Db 542 CTTGGGCATGACGATGGGTGGCAGACAACTTGCCTGGGTTCCGGCGGACCGCAAGGT 601
 QY 599 GACATCTTTGGCGAGGCGGCGAGCATGTTCGTTGTGGCCACCTCATCTGGAAACGACGGC 658
 Db 602 GACATCTTTGGGAGGCGGCGAGCATGTTCGTTGTGGCCACCTCTCTGGAATGGCGGC 661
 QY 659 GACAAACAGGTACAAAGGCGGCGGTTGTTCCGCGCGGCGCATCATGAGGAGGACCATGGTG 718
 Db 662 GACAAACAGGTACAAAGGCGGCGGTTGTTCCGCGCGGCGCATCATGAGGAGGACCATGGTG 721
 QY 719 CCGGACCCCGGTGGACGCGCATGACGCAACAGATCTACGACCTCTTTGTCTCGAGTGT 778
 Db 722 CCGGACCCCGGTGGACGCGCATGACGCAACCTATGG-ACCAAATCTATGACAGTTGGTGGC---TACG 777
 QY 779 GGCTGTGGAGCGCGAGCGAGCATGCTGGTGTGGCGAGTGGGAGTGGAGCACCTTGTCTC 838
 Db 778 GGCTGTGGAGCGCGAGCGAGCATGCTGGTGTGGCGAGTGGGAGTGGAGCACCTTGTCTC 837
 QY 839 GATGCCACCAACAACTCTCTGGGTTCTTGGCGTCTCTCTCTCTCTCTCTCTCTCTCTCT 898
 Db 838 GATGCCACCAACAACTCTCTGGGTTCTTGGCGTCTCTCTCTCTCTCTCTCTCTCTCTCT 897
 QY 899 CCGGACCGCAAGAACATCACCGATGATGTACAAAGTTGGTGGCGGACCGCAAGTATGCA 958
 Db 898 CCGGACCGCGCCAACTCACCGATGATGTACAAAGTTGGTACGCGACGCGCAAGTATGCA 957
 QY 959 AGCGTTCCCGTGTATGATGGGACCAAGACGAGGAGGACCATCTTTTGGCTCTTGAACG 1018
 Db 958 AGCGTTCCCGTGTATGATGGGACCAAGACGAGGAGGAGTCTTTTGTCTCTTGAACA 1017
 QY 1019 TGACCAAGCATGCTCAAGGCGCGTCTTACTTCAAGCAGTTTCAATCCACGCGGACGCGCG 1078
 Db 1018 CCACCAAGCGGCGGCGGAGGAGTCTCTCAAGAGTTTCAATCCACGCGGACGCGCG 1077
 QY 1079 AGATGACACCTTTGATGGCGGCGTACCCCGAGGACATCACCCAGGTTCCGTTGACACGG 1138
 Db 1078 ATATCACCGCATTTGAAGCGGCGTACCCAGCGATGTACCCAGGTTCCGTTGACACGG 1137
 QY 1139 -----TTCAACGCTCACCGCGCATTTCAAGAGATC---GCGGTCTCGCGACCTTG 1188
 Db 1138 GCATTTCAACGCGCTTACACCCAGCTCAAGCGGATCAATGTGTGTGTGGCGACCTCA 1197
 QY 1189 CATTATCCAGCGCGCGCTACTTCTCAACCATCTCCAGGCGGACCAAGTACTCTGT 1248
 Db 1198 CTTTACCTCTCGCGCGCTACTTCTCAACCATCTCCAGGTTCCGTTGACGATCTGT 1257
 QY 1249 TCCTCAAGCAGCTCGGTTGCCAATCATGGGCACTTCCATGCCAACGACATTTGTGTGGC 1308
 Db 1258 TCCTCAAGCAGCTTGGTTGCCCATCTCGG--ACGTTCCACGCGAAGACATTTGTGTGGC 1315
 QY 1309 AGGACTACTTTGGGAAGCGGAGGCGTATCTAACAACGGTTTATTCGGTTCGCCA 1368
 Db 1316 AGGACTTTTGTGGGAGCGGCGTATCTTACAAACCGGTTTCAATCGGTTTGGCA 1375
 QY 1369 CGGACTTGACCCCAACACCGCGGCGTGTGTGTGAATCGGCCAAGT--ACACGACGAG 1426
 Db 1376 CGGACTTGACCCCAACACCGCGG---CTTGTGAGTGGGCCAAGTCAACGACGAGCA 1432
 QY 1427 CCAGGCGCAACAACTTGATGATCAACGCTTGGGCTTGTACACCGGCAAGGACAACTT 1486
 Db 1433 GCGGCGGAGCAACTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1492
 QY 1487 CCGGACCGGTGGTACGCGGTTGATGACCAACCGGTTCTTTGT 1531
 Db 1493 CCGGACCGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1537

X16712
 VERSION X16712.1 GI:2548
 KEYWORDS lipase; lipase 1.
 SOURCE Candida cylindracea
 ORGANISM Candida cylindracea
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; mitosporic Saccharomycetales; Candida.
 REFERENCE 1 (bases 1 to 1657)
 AUTHORS Kawaguchi, Y., Honda, H., Taniguchi-Morimura, J. and Iwasaki, S.
 TITLE The codon CUG is read as serine in an asporogenic yeast *Candida cylindracea*
 JOURNAL Nature 341 (6238), 164-166 (1989)
 MEDLINE 89384874
 PUBMED 2506450
 FEATURES
 Location/Qualifiers
 1..1657
 /organism="Candida cylindracea"
 /mol_type="mRNA"
 /strain="MS-5"
 /db_xref="taxon:44322"
 /clone="lambda CL115"
 <1..1657
 /note="for lipase 1"
 <1..1617
 /note="Protein sequence is in conflict with the conceptual translation"
 /codon_start=1
 /codon=(seq="ctg",aa:Ser)
 /product="lipase 1"
 /protein_id="CAA34684.1"
 /db_xref="GI:2689215"
 /db_xref="GOA:P20261"
 /db_xref="SWISS-PROT:P20261"
 /translation="SVAAATATLANTGNTIGLNAINEAFILGIPPEPPVGNLRFPD
 PVPYSGSDGQKFTSYSCMQQPEGYEENLPKALDLVMQSKVFEAVSPSEDC
 TINVRPPTAGANLPMVLI FGGGEVGGTSTFPQAQMTKSIAMGKPIHVSVMY
 RVSWGFLAGDIKAEGSANAGLQDRLGMQWADNAFGGDPKVTIFGEGSWS
 VMCHLWNGDNTYKGPLFRAGIMOSGAMVPSDAVDGIYGNIEFDLLASNAGCSAS
 DKLACLRSSDNTLEDATNTNTPGLAYSLSLSLPRPDGNIITDDMYALVREGKYN
 IPVIGDQNDGTFPGTSSLSNVTDAQAREYFKQSFOHASDAEIDTLMTAYPDITQG
 SPFDGILNALTPQKRSIVAGLDGLFTLARRYFLNHYTGKTKYSLKSLGSLFVLG
 TFSNDIIVFDLGLSGSLIYNNAFIAFADLPNTAGLLVKKWPEYTSOSSGNLMM
 INALGLYTKDNFRTAGYDALFSNPPSPFV"
 <1..12
 sig_peptide
 mat_peptide
 13..1614
 /product="lipase 1"
 ORIGIN
 Query Match 66.9%; Score 1025; DB 8; Length 1657;
 Best Local Similarity 84.7%; Pred. No. 2.4e-159;
 Matches 1342; Conservative 0; Mismatches 180; Indels 62; Gaps 15;
 QY 2 CCCACCGCAAGCTCGCCAAACGGGACACCATCACCGGTCTCAACGCCATCATCAACGAG 61
 DB 16 CCCACCGCAAGCTCGCCAAACGGGACACCATCACCGGTCTCAACGCCATCATCAACGAG 75
 QY 62 GGGTTCCTCGGATTCCTTTCGCGAGCGCGGCGGCAACCTCCGCTTCAAGACCT 121
 DB 76 GGGTTCCTCGGATTCCTTTCGCGAGCGCGGCGGCAACCTCCGCTTCAAGACCT 135
 QY 122 GTGCGGTACTCTGGCTCGCTCAACGGCCAGAAAGTT-----ACTTACGGCCCGGTCATG 174
 DB 136 GTGCGGTACTCTGGCTCGCTCGATGGCCAGAAAGTTCAAGCTGACGGCCGCTGTGCATG 195
 QY 175 CAGCAGAACCCCGAGGGGACGTTTGAAGAGAACCTTGGCAGACGGCACTCGACTGGTG 234
 DB 196 CAGCAGAACCCCGAGGGGACCTTACGAGGAGAACCTCCCGCAGGACGGCTCGACTGGTG 255
 QY 235 ATGCAAGTCCAGGTTTCCAGCGGTGCTTCCCGAGAGTGGAGTGGCTCTCAACATCAAC 294
 DB 236 ATGCAAGTCCAGGTTTCCAGCGGTGCTTCCCGAGAGTGGAGTGGCTCTCAACATCAAC 315
 QY 295 GTGGTGGCGCGCGGCGCAACGAGCGCGGCGCAACCTCCGGTCAATGCTCTGGATCTTTT 354

Db 316 GTGGTGGCGCGCGCGGACCAAGCGGGTGCACCAACCTCCCGGTGATGCTCTGGATCTTT 375
 QY 355 GCGGTGGTGGTGGATCGGACGCCACCATCTTCTCCCGCCAGATGGTCAACGAG 414
 DB 376 GCGCGCGGGTGGAGTGGGGTGGGACGACACCTTCTCCCGCCAGATGATCAACGAG 435
 QY 415 AGTGTGCTCATGGGCAAG-CCATCATCCAGTGGCGGTCAACTACCGTGTGGCTCTGG 473
 DB 436 AGCATTCGCCATGGCAAGCCCATCATCCAGTGGCGGTCAACTACCGGTGTGGCTGG 495
 QY 474 GGGTCTTGGTGGTGGATGATCAAGCGCGAGGGGAGCGGGAACCGCGGTGGAGGAC 533
 DB 496 GGGTCTTGGTGGCGACGAGATCAAGCGCGAGGGGAGTGCCAAACGCGGTGGAGGAC 555
 QY 534 CAGCGTGGTGGCATGGTGGCGAGACAACATTTGCGGGTTCGGCGGGGACCCGAGC 593
 DB 556 CAGCGTGGGCGATGGAGTGGGCGGACAAATTTGGGGTGGGGGGGACCCGACC 615
 QY 594 AAGGTGA-CATCTTTGGCGA---GGCGGGCAGCATGTCGGTGTGTGCGCACTCATCTGG 649
 DB 616 AAGGTGACCATCTTTGGCGAGCTGGCGGGCAGCATGTCGGTCAATGTGCCACATCTCTGG 675
 QY 650 AACGACGGCGACACAGTACAGGGGCAAGCGGTGTTCGCGGGGCGGATCATGCA---G 706
 DB 676 AACGACGGCGACACAGTACAGGGGCAAGCGGTCTTCCGCGGGGCGGATCATGCGGTG 735
 QY 707 GGAGCCATGGTGGC---GGACCCGTGGACGGCGAGTACGCAACGAGATCTACGACCTC 763
 DB 736 GGGGCCATATGTTGGCGCTGGAGCGCGGTGACCGGATCTACGGCAGAGATCTTGGACCTC 795
 QY 764 TTTGCTCGAGTGGTGGTGGCGGCGCCAGCGCAAGCTCGGTGTGTGCGCAGTGC- 822
 DB 796 TTGGGTGCGACGCGGGCTGGCGAGCGCGCAGCGCAAGCTTGGTGTGTGCGCGGTGTG 855
 QY 823 --GAGCGACACCTTGTGTCGATGCCACCAACAACTCTCTGGTCTTGGCGTACTCTCG 880
 DB 856 CTGAGCGACACGTTGGAGGACCGCCAAACCAACCCCTGGGTGTTCTGGCGTACTCTCG 915
 QY 881 TTGGCGGTG-----TACTCCCGCGCGACGGCAAGACATCACCGGATGATGTACAG 934
 DB 916 TTGGGTGCTGTGTAACCTCCCGCGCGCGCGGTGAACATCACCGACGATGTACGCC 975
 QY 935 TTGGTGGCGGACGCGCAAGTATGCAAGCGTTCCTGTGATCATTTGGCGACAGAACGACG 994
 DB 976 TTGGTGGCGGAGGCAAGTATGCAACATCTCTGTGATCATCTGGCGGACAGAACGACG 1035
 QY 995 GGCACCATCTTTG-----GCTCTTGAACGTGACCAAGATGCTCAGGCGCGTCTTAC 1047
 DB 1036 GGCACCTCTTTTGGCACCTCTGTGTGAACGTGACCAAGATGCCAGCGCGCGGAGTAC 1095
 QY 1048 TTCAAGCA---GTTTCATCCAGCGACGCGGAGATCGACACCTTGTATGGCGCGGTAC 1104
 DB 1096 TTCAAGCAAGTGTGTGTCCAGCGACGCGCGGAGATCGACAGTGTATGACGCGGTAC 1155
 QY 1105 CCCAGGACATCACCCAGGG---TCCGTTTCAGACGCGTT-----CAACGCTCACCCG 1154
 DB 1156 CCNNNGACATCACCCAGGGCTGCGGTTCGACACGCGGTATTTCTCAACGCGCTCACCCG 1215
 QY 1155 CAGTTCAAGAGATC---GCGGTGCTGGCGACCTTGCATTCATCCAGCGCGCGGTAC 1211
 DB 1216 CAGTTCAAGAGAACTCTGGCGGTGCTGGCGACCTTGGCTTTACGCTTGTCTGCTGCTAC 1275
 QY 1212 TTCTCTCAACCTTCCAGGGCGGACCAAGTACTCTGTTCTCTC-----AAGCAGCTCGG 1265
 DB 1276 TTCTCTCAACCTACACCGGGCGGACCAAGTACTCTCTCTCTCTCTCTCTCTCTCTCT 1335
 QY 1266 TTGCAATCATGGGCACTTCCATGCGCAACGACATTTGTGTGGGAGGACTACTTGTGGGA 1325
 DB 1336 TTGCGGTGCTCGGAACGTTTCACTCCAACGACATTTGTCTCAGGACTACTTGTGGGC 1395
 QY 1326 AGCGCAGCGGTCTATCAACAAACGCGTTTATCGGTTTCGCCACGACTTGGACCCCAAC 1385

Db 1396 AGCGGCTCGCTCATCTACAAACACGCGTTCAATGCGTTTGCCACGACATTGGACCCCAAC 1455
QY 1386 ACCGCGGGGTGTGTGTAAGTGGCCCAAGTACACGAGCAGC-----CAGGSCAACAAAC 1439
Db 1456 ACCGCGGGGTGTGTGTAAGTGGCCCGAGTACACGAGCAGCTGAGCTGGGCAACAAAC 1515
QY 1440 TTGATGATGATCAACCGCCTTTGGGCTTTGTACACCGGCAAGGACAACTTCGCGACCGTGGC 1499
Db 1516 TTGATGATGATCAACCGCCTTTGGGCTTTGTACACCGGCAAGGACAACTTCGCGACCGCGGC 1575
QY 1500 TAGGACGCTTGTATGACCAACCG 1523
Db 1576 TAGGACGCTTGTATGACCAACCG 1599

RESULT 6
A48377
LOCUS A48377 1733 bp DNA linear PAT 07-MAR-1997
DEFINITION Sequence 6 from Patent WO9603511.
ACCESSION A48377
VERSION A48377.1 GI:2302168
KEYWORDS
SOURCE Candida rugosa
ORGANISM Candida rugosa
Bukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 1733)
AUTHORS Alibert G., Mouloungui, Z. and Boudet, A.
TITLE METHOD FOR PRODUCING FATTY ACIDS OR DERIVATIVES THEREOF FROM OIL
PLANT
JOURNAL Patent: WO 9603511-A 6 08-FEB-1996;
TOULOUSE INST NAT POLYTECH (FR)
COMMENT Other publication AU 2984995 960222
Other publication FR 2722798 960126.
FEATURES
source Location/Qualifiers
1..1733
/organism="Candida rugosa"
/mol_type="unassigned DNA"
/db_xref="taxon:5481"

ORIGIN
Query Match 66.7%; Score 1022; DB 6; Length 1733;
Best Local Similarity 84.7%; Pred. No. 7.4e-159;
Matches 1342; Conservative 0; Mismatches 180; Indels 62; Gaps 15;
QY 2 CCACCGCAAGCTCGCCACGGGACACCATCACCGTCTCAACGCCATCATCAACGAG 61
Db 49 CCACCGCACGCTCGCCACGGGACACCATCACCGTCTCAACGCCATCATCAACGAG 108
QY 62 GCGTTCTCGGCATTCCCTTTGCGAGCGCGCGTGGCAACCTCGCGCTTCAAGGACCCCT 121
Db 109 GCGTTCTCGGCATTCCCTTTGCGAGCGCGCGTGGCAACCTCGCGCTTCAAGGACCC 168
QY 122 GTGCGGTACTGTGGTCTGCTCAACGGCGAGAGTT-----ACTTACGGCCCGTGATG 174
Db 169 GTGCGGTACTGTGGTCTGCTCGATGCGCAGAGTTACCGCTGACGCGCGCGTGTGATG 228
QY 175 CAGCAACACCCGAGGCGAGTTTGAAGAGAACTTGGCAAGACGCACTCGACTTGGTG 234
Db 229 CAGCAACACCCGAGGCGACCTACGAGGAGAACTTCCCAAGGACGCGCTCGACTTGGTG 288
QY 235 ATGCACTCAAGGTGTTCACGGCGGTGTTCCCGCAGAGTGAGGACTGCGCTCACCATCAAC 294
Db 289 ATGCACTCAAGGTGTTCGAGGCGGTGCTCGCGCTGAGCGAGACTGTCTCACCATCAAC 348
QY 295 GTGGTGGCGCGCGGACCAAGGGGCGCAACCTCCCGGTATGCTCTGATCTTT 354
Db 349 GTGGTGGCGCGCGGACCAAGGGGCGCAACCTCCCGGTATGCTCTGATCTTT 408
QY 355 GGCGGTGGGTTCAGATCGGCAGCCCAACCATCTTCCTCCCGCCAGATGGTCAACAG 414
Db 409 GGCGGTGGGTTCAGATCGGCAGCCCAACCATCTTCCTCCCGCCAGATGGTCAACAG 468

QY 415 AGTGTCTCATGGGCAAG--CCATCATCCAGTGGCGGTCAACATACCGTGTGTGCTCGTGG 473
Db 469 AGCATTTGCCATGGCAAGCCCATCATCCAGTGGAGGTCAACTACCGGTGTGTGCTGG 528
QY 474 GGGTTCTTGGCTGTGTATGACATCAAGGCGGAGGAGGAGGCAACCGCGCTTTGAAGGAC 533
Db 529 GGGTTCTTGGCTGGCGACGAGATCAAGGCGGAGGAGTCCCAACCGCGTTTGAAGGAC 588
QY 534 CAGCGTTTGGGATGCAAGTGGGTGGCAGACAACATTGCCGGTTCGGCGCGGACCCGAGC 593
Db 589 CAGCGTTTGGGATGCAAGTGGGTGGCGACAACATTGCCGGTTCGGCGCGGACCCGACC 648
QY 594 AAGGTGA-CATCTTTGGCGA---GGCGGCGAGCATGTCCGTGTGTGGCACTCATCTGG 649
Db 649 AAGGTGACCATCTTTGGCGAGCTGGCGGCGAGCATGTGGTGTATGTGCACATCTCTGG 708
QY 650 AAGCAGCGGCAACACAGTACAAGGCAAGCGTGTTCGGCGCGGCGCATCATGCA---G 706
Db 709 AAGCAGCGGCAACACAGTACAAGGCAAGCGCTCTTCGGCGCGGCGCATCATGCAAGCTG 768
QY 707 GGAGCCATGTGCC---GGACCCGCTGGAGCGGCACTAGCGCAACGAGATCTACGACCTC 763
Db 769 GGGGCCATGTGCGCGTGGACGCGGTGGAGCGCATCTACGGCAACGAGATCTTTGACCTC 828
QY 764 TTTGTCTGAGTGTCTGGCTGTGGCAGCGCAGCAAGCTCGCGTGTTCGGCAGTGC- 822
Db 829 TTGGCGTTCGAACGCGGCTGGCGACGCGCCAGCAAGCTTGGCTGTTCGGCGGTGTG 888
QY 823 --GAGGCAACCTTGTCTGATGCGCAACACACACTCTCTGGTTCCTGGGTACTCTCTG 880
Db 889 CTGAGCGACACGTTGGAGGACGCCAACAAACACCCCTTGGGTCTTTCGGCTACTCTCTG 948
QY 881 TTGGCGTTCG---TACTCCCGCGCGCGGCAAGCAACATCACCGATGACATGTACAAG 934
Db 949 TTGGCGTTCGTTACTCTCCCGCGCGCGGCGGAGATCACCGCAACGATGTAGGCC 1008
QY 935 TTGGTGGCGGACGGCAAGTATGCAAGCTTCCCGTGTATCATTTGGCGACAGCAACGAG 994
Db 1009 TTGGTGGCGGAGGCAAGTATGCAACATCCCTGTGTATCATCGCGACAGCAACGAG 1068
QY 995 GGCACCATCTTTG-----GCTTTGAACGTGACCAAGATGCTCAGGCGCGTGTCTAC 1047
Db 1069 GGCACCTTCTTTGGCACCCCTGTGTGTAAGCTGACCAACGAGTCCCGAGGCGCGAGTAC 1128
QY 1048 TTCAAGCA--GTTTATCCAGCGCAGCGAGAGATGACACCTTGTATGGCGGTAC 1104
Db 1129 TTCAAGCACTGTTTGTCCAGCGCAGCGAGAGATGACACGTTGATGAGCGGTAC 1188
QY 1105 CCCAGGACATCACCCAGGG---TCCGTTTCGACACGGTT-----CAAAGCTCACCCCG 1154
Db 1189 CCGCGGACATCACCCAGGCGCTGCGTTTCGACACGGGTATTCTCAACGCGCTCACCCCG 1248
QY 1155 CAGTTCAAGAGATC---GCGGTGCTGGGAGCTTGCATTCATCCAGCGCGCGCGGTAC 1211
Db 1249 CAGTTCAAGAGATTCCTGGCGGTGCTGGCGACCTTGGCTTTACGTTGCTGCTCGGTAC 1308
QY 1212 TTCTCAACCACTTCAGGCGGCGACCAAGTACTCGTTCTCTC-----AAGCAGCTCGGG 1265
Db 1309 TTCTCAACCACTACACCGCGGCGACCAAGTACTCATTCCTCTGAAGCAGCTCTCGGGC 1368
QY 1266 TTGCCAATCATGGGACCTTTCATGCAACGACATTTGTGTGGCAGGACTACTTTTGGGA 1325
Db 1369 TTGGCGGTCTCGGAACGTTTCACTTCAACGACATTTGTCTTCAGGACTACTTTTGGGC 1428
QY 1326 AGCGGAGCGGTCTATCAACAAACGCGTTTATCGCGTTCGCAACGAGCTTGAACCCCAAC 1385
Db 1429 AGCGGTCTCATCTACAAACGCGTTCAITTCGCGTTTTCGCAACGAGCTTGAACCCCAAC 1488
QY 1386 ACCGCGGGGTGTGTGTGAACCTGGCCCAAGTACACCGAGC-----CAGGCAACAAAC 1439
Db 1489 ACCGCGGGGTGTGTGTGAAGTGGCCCGGAGTACACCGAGCGCTGAGCTGGGCAACAAAC 1548
QY 1440 TTGATGATGATCAACCGCTTGGGCTTGTATACCGGCAAGGACAACTTTCGCGACCGGTGGC 1499

```
Db 1549 TTGATGATCAACGCTTGGGCTTGATACCGGCAAGGACAATCTCCGACCGCGC 1608
|||||
Qy 1500 TACGACGGTTGATGACCAACCCG 1523
|||||
Db 1609 TACGACGGTTGTTCTTCCAACCCG 1632
|||||

RESULT 7
CCLIP1 1733 bp DNA linear PLN 27-APR-1993
LOCUS C.cylindracea Lip1 gene for lipase.
DEFINITION X64703
ACCESSION X64703.1 GI:2544
VERSION Lip1 gene.
KEYWORDS Candida cylindracea
SOURCE Candida cylindracea
ORGANISM Candida cylindracea
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 1733)
AUTHORS Longhi S., Fusetti, F., Grandori, R., Lotti, M., Vanoni, M. and
Alberghina, L.
TITLE Cloning and nucleotide sequences of two lipase genes from Candida
cylindracea
JOURNAL Biochim. Biophys. Acta 1131 (2), 227-232 (1992)
MEDLINE 92305068
PUBMED 1610906
REFERENCE 2 (bases 1 to 1733)
AUTHORS Longhi, S.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-1992) S. Longhi, Universita' degli Studi di
Milano, Via Celozia 26, 20133 Milano, ITALY
COMMENT See also X64704
X66006, X66007 and X66008 are related sequences in the description
of C.cylindracea Lip1 and Lip2 sequences (X64703, X64704).
FEATURES
source
1..1733
/organism="Candida cylindracea"
/mol_type="genomic DNA"
/strain="ATCC 14830"
/db_xref="taxon:44322"
/clone="A"
1..1650
/gene="LIP1"
1..1650
/gene="LIP1"
/EC_number="3.1.1.3"
/codon_start=1
/product="triacylglycerol lipase"
/protein_id="CAA45957.1"
/db_xref="GI:1325987"
/db_xref="GOA:P20261"
/db_xref="SWISS-PROT:P20261"
/translation="METALALLIASVAAPATLAIANGDTTGLNAINAEFLGIPFA
EPVGNLRFKDPVPYSGSLDQKFTLYGLPCMQNPEGYEENLPKALDILVMSKVF
EAVPLSDCLTINVVRPGTKAGANLFWLIFGGFVGSTSTFPDPAQMITKSIAM
KPIITHSVNRYKSWGLAGDEIKAEKSNAGLQDRLGOWADNIAAFGGDPTKV
TIFGLAGSMVCHILWANDGNTYKGLPFRAGIMOLGANVPLDADVGIYGNBIFDL
LASNAGCSADKLAICRGVLSLTLEDATNTPGFLAYSSLRLLYLPRPDGWNIDDM
YALVREGKIANIPIVIGQNDGFTFGTLLNVTIDQAREFFKQLFVHSDAIEDTL
MTAYPGDITQGLPFDPTGILNALTPOFKRILAVIGDLFTLARRVPLNHYTGKYSFL
LKQLGLPVLGTFHSDINIVFDYLLIGSGLIYNNAFIAFATDLDPNTAGLLVWFPEYT
SSQLGNLMMINLGLVTGKDNFRTAGYDALFSNPPLFFV"
1..45
sig_peptide
/gene="LIP1"
/mat_peptide
/feature="triacylglycerol lipase"
46..1647
/gene="LIP1"
/product="triacylglycerol lipase"
/EC_number="3.1.1.3"

ORIGIN
Query Match 66.7%; Score 1022; DB 8; Length 1733;
|||||
```

```
Best Local Similarity 84.7%; Pred. No. 7.4e-159;
Matches 1342; Conservative 0; Mismatches 180; Indels 62; Gaps 15;

Qy 2 CCCACCCGCAAGCTCGCCCAAGCGGACACCATCACCGGTCTCAAGCCATCATCAACGAG 61
Db 49 CCCACCCGCAAGCTCGCCCAAGCGGACACCATCACCGGTCTCAAGCCATCATCAACGAG 108
|||||
Qy 62 GCGTTCCTCGGCATTTCCTTTGCCGAGCCGCGGTGGGCAACCTCCGCTTCAAGAGACCT 121
Db 109 GCGTTCCTCGGCATTTCCTTTGCCGAGCCGCGGTGGGCAACCTCCGCTTCAAGAGACCT 168
|||||
Qy 122 GTCCGCTACTCTGCTCGCTCAACCGGCCAGAGTT-----ACTTACGGCCCGGTGATG 174
Db 169 GTCCGCTACTCTCGCTCGCTCGATGGCGCAGAGTTTCACGCTGTACGGCCCGCTGTGCATG 228
|||||
Qy 175 CAGCAGAACCCCGAGGGCAGCTTTTGAAGAGAACCTTGGCAAGCGGCACTCGACTTGGTG 234
Db 229 CAGCAGAACCCCGAGGGCAGCTTGAAGAGAACCTTGGCAAGCGGCACTCGACTTGGTG 288
|||||
Qy 235 ATGCAGTCCCAAGGTGTTCCAGGCGGTGTTCCCGCAGAGTGAAGACTGCCTCACCATCAAC 294
Db 289 ATGCAGTCCCAAGGTGTTTGAAGCGGTGCTGCCGCTGAGCGAGGACTGTCTCACCATCAAC 348
|||||
Qy 295 GTGGTGGCGCCCGCGGACCAAGCGCGGCGGCAACCTCCCGCTCATGCTCTGATCTTT 354
Db 349 GTGGTGGCGCCCGCGGACCAAGCGCGGTGCCAACCTCCCGGTGATGCTCTGGATCTTT 408
|||||
Qy 355 GGCGGTGGTGGTGGATGCGGAGCCCAACCATCTTCCCTCCCGCCAGATGGTCAACAG 414
Db 409 GGCGGTGGTGGTGGATGCGGAGCCCAACCATCTTCCCTCCCGCCAGATGGTCAACAG 468
|||||
Qy 415 AGTGTGCTCATGGGCAAG-COATCATCAAGTGGCGGTCAACTACCGTGTGCTCTGCTG 473
Db 469 AGCATTGCCATGGGCAAGCCCATCATCAAGTGGCGGTCAACTACCGTGTGCTCTGCTG 528
|||||
Qy 474 GGGTCTTGGTGGTGGATGATCAAGSCCGAGGCGGAGGAGCGCGGTGGAGAC 533
Db 529 GGGTCTTGGTGGCGAGAGATCAAGSCCGAGGCGGAGTGGCAACCGCGGTGGAGAC 588
|||||
Qy 534 CAGCGTGGTGGCGATGCTGGTGGCGAGCAACAATGGCGGTGGCGGCGGACCCGAGC 593
Db 589 CAGCGTGGTGGCGATGCTGGTGGCGAGCAACAATGGCGGTGGCGGCGGACCCGAGC 648
|||||
Qy 594 AAGGTGA-CATCTTTGGCGA---GGCGGGCAGCATGTCCTGTTGTGTGCCACTATCTGG 649
Db 649 AAGGTGACCATCTTTGGCGAGCTGGCGGGCAGCATGTCCTGTTGTGTGCCACTATCTGG 708
|||||
Qy 650 AACGACGGCGAACAACGTAAGGGCAAGCGGTGTTCCGCGCGGCGCATATGCA--G 706
Db 709 AACGACGGCGAACAACGTAAGGGCAAGCGGTGTTCCGCGCGGCGCATATGCAAGTGG 768
|||||
Qy 707 GGAGCCATGTTGCC---GGACCCCGTGGACGCGCATGTCAGGCAACGAGATCTTACGACCTC 763
Db 769 GGAGCCATGTTGCCCGCTGGAGCGCGTGGACGCGCATGTCAGGCAACGAGATCTTACCTC 828
|||||
Qy 764 TTTGTCTCAGTGTGGTGTGGCAGCGCCAGCGAAGTTCGCGTGTGTCGGCAGTGC- 822
Db 829 TTGGCGTGAACCGCGGGCTGGCGGAGCGCGCAGCGCAAGTTCGCGTGTGTCGGCGGTGTG 888
|||||
Qy 823 --GAGCGACACTTGTCTCGATGGCCACCAACAACCTCCGCGGTGTTCTGCGGTACTCTCG 880
Db 889 CTGAGCGACACTTGTGGAGGACGCCACCAACAACCTCCGCGGTGTTCTGCGGTACTCTCG 948
|||||
Qy 891 TTGGCGTTG-----TACTCCCGCGCCGACGCGCAAGAACATCACCAGTGAATGTACAG 934
Db 949 TTGGCGTTGCTGTACTCTCCCGCGCCGCGCGGTGAACATCACCAGGACATGTACGCC 1008
|||||
Qy 935 TTGGTGGCGAGCGGCAAGTATGCAAGCGTTCCGCTGATCATTTGGCGACAGAACGAG 994
Db 1009 TTGGTGGCGAGCGGCAAGTATGCAAGCATCCCTGTGATCATGGCGGACAGAACGAG 1068
|||||
Qy 995 GGCACCATCTTTG-----GCTCTTGAACGTGACCAAGATGTCTCAGCGCCGCTGTAC 1047
|||||
```

```

Db      1069 GGCACCTTTCTTTGGCACCCCTGCTGTTGAAGTGAACCGATGCCAGGCGCCGCGAGTAC 1128
Qy      1048 TTCAAGCA---GTTTATCCAGCGCAGCGACGCGAGATGACACCTTGATGGCGCGTAC 1104
Db      1129 TTCAAGCAGCTGTTTGTTCAGCCAGCGACGCGAGATGACACCTTGATGACGCGTAC 1188
Qy      1105 CCCAGGACATCAACCCAGGG---TCCGTTGACACGGTT-----CAAAGCTCACCCCG 1154
Db      1189 CCGGCGACATCACCCAGGCGCTGCCGTTGCACACGGGTATTCTCAAGCGCCTCACCCCG 1248
Qy      1155 CAGTTCAAGAGATC---GGGTTGCTCGGCGACCTTGATTCATCCAGGCGCGGTAC 1211
Db      1249 CAGTTCAAGAGATCCTCGGCGGTGCTCGGCGACCTTGCTTACGCTTGCTCGCGTAC 1308
Qy      1212 TTCTCAACCACTTCCAGGCGCGCACCAAGTACTCGTTCTCTC-----AAGCACTCGGG 1265
Db      1309 TTCTCAACCACTACACGCGCGGACCAAGTACTCATCTCTCTGAAGAGCTCCTGCGC 1368
Qy      1266 TTGCCAATCATGGGCACTTCCATGCGCAAGACATTTGTGGCAGGACTACTTGTGGGA 1325
Db      1369 TTGCGCGTGTCTGGAACGTTTCCATCCCAACGACATTTGCTTCCAGGACTACTTGTGGC 1428
Qy      1326 AGCGGCGAGCTCATCTACAAACGCGTTTATCGCTTGGCACCAGCTTGGAACCCCAAC 1385
Db      1429 AGCGGCTCGCTCATCTACAAACGCGTTTATTTGCGTTTGGCAGGACTTGGAACCCCAAC 1488
Qy      1386 ACCGCGGGGTTGTTGTGTAAGTGGCCCAAGTACACCAAGCAGC-----CAGGGCAACAA 1439
Db      1489 ACCGCGGGTGTGTTGTAAGTGGCCGAGTACACCAAGAGCTGCGAGCTGGGCAACAA 1548
Qy      1440 TTGATGATGATCAACGCTTGGGCTTGTAACCGGCAAGGACAACTTCCGACCGCTGGC 1499
Db      1549 TTGATGATGATCAACGCTTGGGCTTGTAACCGGCAAGGACAACTTCCGACCGCGGC 1608
Qy      1500 TAGCAGCGTTGATGACCAACCCG 1523
Db      1609 TAGCAGCGTTGTTCTCCAACCCG 1632

RESULT 8
AX709927
LOCUS      AX709927          1511 bp      DNA      linear      PAT 10-APR-2003
DEFINITION Sequence 9 from Patent EP1288294.
ACCESSION  AX709927
VERSION     AX709927.1  GI:29786305
KEYWORDS
SOURCE      Candida rugosa
ORGANISM    Candida rugosa
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE   1
AUTHORS    Tang, S.J., Lee, G.C. and Shaw, J.F.
TITLE       Recombinant candida rugosa lipases
JOURNAL     Patent: EP 1288294-A 9 05-MAR-2003;
            Academia Sinica (TW)
FEATURES
    source   1..1511
             /organism="Candida rugosa"
             /mol_type="unassigned DNA"
             /db_xref="taxon:5481"

ORIGIN
Query Match      65.8%; Score 1007.6; DB 6; Length 1511;
Best Local Similarity 85.7%; Pred. No. 1.8e-156;
Matches 1313; Conservative 0; Mismatches 184; Indels 35; Gaps 16;

Qy      3  CCACCGCAAGCTCGGCAACGGCGACACCATCCCGTCTCAACGCCATCATCAACGAGG 62
Db      2  CCACCGCAAGCTCGGCAACGGCGACACCATCCCGTCTCAACGCCATCATCAACGAGG 61
Qy      63  CGTTCTCGCATTCCTTTGCGGAGCGCGGTTGCGGACCTCCGTTCAAGGACCCCTG 122
Db      62  CGTTCTCGCATTCCTTTGCGGAGCGCGGTTGCGGACCTCCGTTCAAGGACCCCG 121

```

```

Qy      123  TGCCTACTCTGGCTCGCTCAACGCGCCAGAGTTTACTTACGCGCCGTCGATGCAGAGAA 182
Db      122  TGCCTACTCTGGCTCGCTCGATGGCCAGAGTTTACTTACGCGCCGTCGATGCAGAGAA 181
Qy      183  CCCGAGGGACGCTTTGAAGAGAACTTTGGCAAGACGGCACTCGACTTGTGTGATGTCAGTC 242
Db      182  CCCGAGGGACCTTACGAGGAGAACTTCCCAAGGACGCGCTCGACTTGTGTGATGTCAGTC 241
Qy      243  CAAGGTGTTCCAGGCGGTGCTTCCCAAGAGTAGGAGTCTGCTTACCATCAACGTGTTGG 302
Db      242  CAAGGTGTTTGAAGCGGTG-----CCGAGCGAGGACTGTCTCACCATCAACGTGTTGG 295
Qy      303  GCGCGCGGGCACCAAGGCGCGCGCAACCTCCGCTCATCTCTGAGTCTTTGGCGGTGG 362
Db      296  GCGCGCGGGCACCAAGGCGGTGCGCAACCTCCCGTGTATCTCTGAGTCTTTGGCGGTGG 355
Qy      363  GTTTGAGATCGGAGCGCCCAACCATCTTCCCTCCCGCCAGATGCTCACCAAGAGTGTGCT 422
Db      356  GTTTGAGGTGGTGGCACCAAGCACTTCCCTCCCGCCAGATGATCACCAAGAGCATGTC 415
Qy      423  CATGGGCAAG-CCATCATCTCAGTGGCGCTCAACTACCGTGTGCTCTCGTGGGTTCTT 481
Db      416  CATGGGCAAGCCCATCATCCACGTGAGCGTCAACTACCGCTGTCTCGTGGGTTCTT 475
Qy      482  GGCTGTGTGATGACATCAAGCGCGAGGCGAGCGGGAACGCGGCTTGAAGGACCAAGCGTTT 541
Db      476  GGCTGGCGCAGATCAAGGCCGAGGCACTGCCAACGCCGTTTGAAGGACCAAGCG-CT 534
Qy      542  GGGCATGCACTGAGTGGTGGCAGACAACATTTGCGGGTTTGGCGGGACCCGAGCAAGTGA- 600
Db      535  GGGCATGCACTGAGTGGTGGCGGACAACATTTGCGGCTTTTGGCGGACCCGACCAAGGTGAC 594
Qy      601  CATCTTTGGCGAGCGCGGAGCATGTCCGTGTTGTGCCACTCATCTGGAACGAGCGGA 660
Db      595  CATCTTTGGCGAGCGCGGAGCATGTCCGTGTTGTGCCACTCATCTGGAACGAGCGGA 654
Qy      661  CAACAGCTACAAGGGCAAGCGCTTGTTCGCGGGGCATCATCGAGGAGCCATGTTGCC 720
Db      655  CAACAGCTACAAGGGCAAGCGCTTTCGCGGGGCATCATCGAGGAGCCATGTTGCC 714
Qy      721  GGAACCGGTGGAGCGGACGCTACGGCAACAGAGATCTACGACCTTCTTCTCGAGTGTGG 780
Db      715  GGAACGC-GTGGACGGCTC-TACGGCAACAGAGATCTTGAACCTTCTGCGCTCGA-CGCGGG 771
Qy      781  CTGTGCGAGCGCGGAGGACGAGCTCGGTGCTTGGCAGTGGCGAGGACACCTTGTCTGA 840
Db      772  CTGCGGAGCGCGGAGGACGAGCTTGGCTGCTTGGCGGTGTGAGCGACACGTTGGAGGA 831
Qy      841  TGCCACCAACACACTCTCTGGGTTCCTGGCTACTCTCTGTTGCGGTGTGTACT-CCGCGC 899
Db      832  CGCCACCAACACACCTCTGGGTTCCTGGCTACTCTCTGTTGCGGTGTGTACTCTCCCGGC 891
Qy      900  CGGACGGCAAGAACATCACCGATGACATGTACAAGTTGGTGGCGGACGCGCAAGTATGCA 959
Db      892  CGGACGGGTGAACATCACCGACGACATGTGC---CTTGTGTCGAGGCGCAAGTATGCA- 947
Qy      960  GCGTTCCCGTGTATCTTGGGACCAAGACGAGGCGGACCATCTTTGGC-TCTTGAACG 1018
Db      948  ---CTCCTGTGTATCATCGGCGACGAGACGAGGCGGACCTTCTTTGGCACCTTGAACG 1004
Qy      1019  TGACCAAGATGCTCAGGCGCGGTCTTACTTCAAGCAGTTTCAACGCGGACGCGGCGG 1078
Db      1005  TGACCAAGATGCGAGGCC---GCAGTACTTCAAGCAGTTTGTTCACGCGGACGCGG 1061
Qy      1079  AGATCGACACCTTGTATGGGCGGTACCCCGAGGACATCACCCAGGTCGCTTCGACACG 1138
Db      1062  AG-TCGACAGTTGATGACGCGGTACCCGACATCACCCAGGCGGCTTCGACACGCGTT 1120
Qy      1139  TTCAAG--CTCACCCCGAGTTCAAGAAATCGCGGTGCTCGGCGACCTTGTGATTCATC 1196
Db      1121  CTCAAGCGCTCAACCCCGAGTTCAAGAAATCGCGGTGCTCGGCGACCTTGTGTTACG-- 1178

```

QY 1197 CAGCCCGCGCTACTTCTCAACCACTTCCAGGCGCGCAACCAAGTACTCTGTTCTCTCAAG 1256
 Db 1179 -----CGTCTACTTCTCAACCACTTCCAGGCGCGCAACCAAGTACTCTCTCAAG 1233
 QY 1257 CAGCTCGGTTGCCAATCATCGGCACTTCCATGCCAAGCAATGTTGGCAGGACTAC 1316
 Db 1234 CAGCTCGGTTGCCGTTCTCGAAGCTTCCACTCCAACGACATTTGTCTTCAGGACTAC 1293
 QY 1317 TTGTTGGAGCGGCGAGGCTCATCTACAAACGCGTTTATCGGTTGCGCACCGACTTG 1376
 Db 1294 TTGTTGGGACGCGCTCGCTCATCTACAAACGCGTTTATCGGTTTGGCCACGACTTG 1353
 QY 1377 GACCCCAACACCGCGGCTTGTGTGAAGTGGCCCAAGTACACAGCAGCAGGCGCAAC 1436
 Db 1354 GACCCCAACACCGCGGCTTGTGTGAAGTGGCCCGAGTACACAGCAGCAGGCGCAAC 1413
 QY 1437 AACTTGATGATCATCAACCGCTTGGCTTGTATACCGGCAAGCAAACTTCGGCACCGCT 1496
 Db 1414 AACTTGATGATCATCAACCGCTTGGCTTGTATACCGGCAAGCAAACTTCGGCACCGCT 1472
 QY 1497 GGCTACGACGCTGATGACCAACCGCTTCTT 1528
 Db 1473 GGCTACGACGCTTGTGTCTTCAACCGCGCTT 1504

RESULT 9
 A81171 LOCUS A81171 1647 bp DNA linear PAT 21-JAN-2000
 DEFINITION Sequence 1 from Patent WO9914338.
 ACCESSION A81171
 VERSION A81171.1 GI:6731630
 KEYWORDS .
 SOURCE unidentified
 ORGANISM unclassified.
 REFERENCE 1 (bases 1 to 1647)
 AUTHORS Schmidt-Dannert,C. and Schmid,R.
 TITLE TOTAL SYNTHESIS AND FUNCTIONAL OVEREXPRESSION OF A CANDIDA RUGOSA
 LIPI GENE CODING FOR A MAJOR INDUSTRIAL LIPASE
 JOURNAL Patent: WO 9914338-A 1 25-MAR-1999;
 SCHMIDT DANNERT CLAUDIA (DE); SCHMID ROLF (DE)
 FEATURES
 source
 1..1647
 /organism="unidentified"
 /mol_type="unassigned DNA"
 /isolate="985229"
 /db_xref="taxon:32644"

ORIGIN
 Query Match 65.4%; Score 1002.4; DB 6; Length 1647;
 Best Local Similarity 84.3%; Pred. No. 1.3e-155;
 Matches 1336; Conservative 0; Mismatches 186; Indels 62; Gaps 16;
 QY 2 CCCACGCCAAGCTCGCCCAACGCGGACCACTACCGGTTCTCAAGCCCATCATCAACGAG 61
 Db 49 CCCACGCCAAGCTCGCCCAACGCGGACCACTACCGGTTCTCAAGCCCATCATCAACGAG 108
 QY 62 GCGTTCTCGGATTCCTTTTCCGAGCGCGGTTGGCAACTCCGCTTCAAGGACCT 121
 Db 109 GCGTTCTCGGATTCCTTTTCCGAGCGCGGTTGGCAACTCCGCTTCAAGGACCT 168
 QY 122 GTGCGCTACTCTGGCTCGCTCAACGCCGAGAAAGTT-----ACTTACGGCCCG--TGCA 174
 Db 169 GTGCGCTACTCTGGCTCGCTCAACGCCGAGAAAGTTCAAGCTTCAAGGACCT 228
 QY 175 CAGCAGAACCCGAGGCGAGTTTGAAGAGAACCTTGGCAAGCAGGCACTCGACTTGGTG 234
 Db 229 CAGCAGAACCCGAGGCGAGTTTGAAGAGAACCTTGGCAAGCAGGCACTCGACTTGGTG 288
 QY 235 ATGCACTTCAAGGTTTCCAGGCGGTTCTTCCAGAGTGGAGTGCCTCACCATCAAC 294
 Db 289 ATGCACTTCAAGGTTTGAAGGCGGTTGAGCCGAGCAGGAGGAGTGTCTCACCATCAAC 348

QY 295 GTGTTGCGGCGCGCGGCGCAACAGGCGGCGCGCAACCTCCCGGTCACTGCTCTGATCTTT 354
 Db 349 GTGTTGCGGCGCGCGGCGCAACAGGCGGCGCGCAACCTCCCGGTGATGCTCTGATCTTT 408
 QY 355 GCGGTTGGTTGAGATCGGCGAGCCCGACCATCTTCCCTCCCGCCAGATCGTCAACCAAG 414
 Db 409 GCGGCGGCGTTGAGGTTGGGTGGCCAGACACCTTCCCTCCCGCCAGATCATCAACCAAG 468
 QY 415 AGTGTGCTCATGGGCAAG--CCATCATCACTGCGCGCTCAACTACCGTTGTCCTCGTGG 473
 Db 469 AGCATTTGGGATGGCAAGCCCATCATCACTGAGCGTCACTACCGGTGTCGTCGTGG 528
 QY 474 GGGTTCTTGGTGTGTGATGATCAAGGCGAGGCGAGCGGGAACCGCGCTTGAAGAC 533
 Db 529 GGGTTCTTGGTGTGTGATGATCAAGGCGAGGCGAGTGCACACCGCGTTGAAGAC 588
 QY 534 CAGGTTTGGCATGAGTGGTGGCGAGCAAACTTCCGCGTTTCGGCGGCGACCGGAGC 593
 Db 589 CAGGCTTGGGATGAGTGGTGGCGAGCAAACTTCCGCGTTTCGGCGGCGACCGGAGC 648
 QY 594 AAGGTGAC-ATCTTTGGCGAG---GCGGCGAGCATGTCCGTTGTGTGCACTCATCTGG 649
 Db 649 AAGGTGAGATCTTTGGCGAGCGCGGCGAGCATGTCCGTTGTGTGCACTCATCTGG 708
 QY 650 AACGACGCGGCAACACAGTACAGGCGAGCGGTTTCCGCGGCGGCGATCATGCGAGC 706
 Db 709 AACGACGCGGCAACACAGTACAGGCGAGCGGCTTTCGCGGCGGCGATCATGCGAGC 768
 QY 707 GGAGCCTATGTTGCGG---GACCGGTGAGCGGCGAGTACGCGCAACAGATCTACGACCTC 763
 Db 769 GGGCGATGTTGCCGAGCGAGCGCGTGGACCGCATCTACGCGCAACGAGATTTTGACCTC 828
 QY 764 TTTGTCTGAGTGTGTTGGCAGCGCCAGCAAGCTCGGTTGTGCGCAGTGGC 823
 Db 829 TTGGGCTGAACGCGGCGTGGCGAGCGCCAGCAAGCTCGCGTGTTCGCGGTGTG 888
 QY 824 ---AGCGACACTTGTCTGATGCGCAGCAACACTCTCTGGTTCTTGGCTACTCTCTCG 880
 Db 889 AGCAGCGACACTTGTGGAGGCGCCACCAACACCCCTGGGTTCTTGGCTACTCTCTCG 948
 QY 881 TTGGGTTGTACT-----CCGCGCCGAGCGCAAGAAACATCACCGATGACATGATCAAG 934
 Db 949 TTGGGTTGAGTACTCTCCCGCGCCGAGCGGCTGAACATCACCGACGATGATCGCC 1008
 QY 935 TTGGTTCGCGAGCGCAAGTATGCAAGCGTTCCGTTGATCATTTGGCGACCGAAGCAGAG 994
 Db 1009 TTGGTTCGCGAGCGCAAGTATGCAAGCATCTCCGTTGATCATTCGCGGACCGAAGCAGAG 1068
 QY 995 GGCACCATCTTTGGC-----TCTTGAACGTGACCAAGAAATGCTCAGGCCGCTGCTTAC 1047
 Db 1069 GGCACCTTCTTTTGGCACCCAGCAGCTTGAACGTTGACCGGATGCGCCAGGCCCGGAGTAT 1128
 QY 1048 TTCAAGCAG---TTTCATCCAGCCGAGCGAGATGACACCTTGTATGCGCGGTAC 1104
 Db 1129 TTCAAGCAGAGTTTGTTCACCGCAGCGAGCGGAGATGACACGTTGATGACGCGGTAC 1188
 QY 1105 CCCCAGGACATCAACCCAGG---GTCGCTTGCACAGGTT-----CAACGCTCACCCCG 1154
 Db 1189 CCGCGGACATCAACCCAGGCGCGCTTCGACACGCGGTATTTCTCAACGCCCTCACCCCG 1248
 QY 1155 CAGTTCAAGAGAAAT---CGCGGTGCTCGCGACCTTGCATTCATCCAGCCCGCGCTAC 1211
 Db 1249 CAGTTCAAGAGAAATCAGCGCGGTGCTCGCGACCTTGGCTTTACGTTGCTGTCGCTAC 1308
 QY 1212 TTCTCTCAACCACTTCCAGGCGCGCAACCAAGTACTGTTCTCTCA-----AGCAGTTCGGG 1265
 Db 1309 TTCTCTCAACCACTTACACCGCGCGCAACCAAGTACTGCTTCTCAGCAAGCAGCTCAGCGGC 1368
 QY 1266 TTGCCAATCATGGGCACTTCCATGCGCAACGAGATTTGTGTGGAGGACTACTTGTGGGA 1325
 Db 1369 TTGGCGGTGCTCGAAGCGTTTCCACTCCAACGAGATTTGCTTCCAGGACTACTTGTGGGC 1428
 QY 1326 AGCGGCGAGCTCATCTACAAACGCGTTTATCGCGTTTCGCCACCGACTTGGACCCCAAC 1385


```
Db 1429 AGCGGCTCGCTCATCTACAAACGCGTTCAITTCGCTTTGCCACGACTTGGACCCCAAC 1488
QY 1386 ACCGCGGGTGTGTGTGTAACCTGGCCCAAGTACACCAAGAGC-----CAGGSCAACAC 1439
Db 1489 ACCGCGGGTGTGTGTGTAAGTGGCCCGAGTACACCAAGAGAGCCAGCAACAC 1548
QY 1440 TTGATGATGATCAACGCTTGGCTTTGTATACCGGCAAGGACAACTTCCGACCCGCTGGC 1499
Db 1549 TTGATGATGATCAACGCTTGGCTTTGTATACCGGCAAGGACAACTTCCGACCCGCGC 1608
QY 1500 TACGAGCGGTTGATGACCAACCG 1523
Db 1609 TACGAGCGGTTGTCTTCCAAACCG 1632

RESULT 10
AX709923
LOCUS AX709923
DEFINITION Sequence 5 from Patent EP1288294.
ACCESSION AX709923
VERSION AX709923.1 GI:29786303
KEYWORDS
SOURCE Candida rugosa
ORGANISM Candida rugosa
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1
AUTHORS Tang, S.J., Lee, G.C. and Shaw, J.P.
TITLE Recombinant candida rugosa lipases
JOURNAL Patent: EP 1288294-A 5 05-MAR-2003;
Academia Sinica (TW)
FEATURES
source
1..1548
/organism="Candida rugosa"
/mol_type="unassigned DNA"
/db_xref="taxon:5491"

ORIGIN
Query Match 65.2%; Score 998.4; DB 6; Length 1548;
Best Local Similarity 83.8%; Pred. No. 5.9e-155;
Matches 1301; Conservative 0; Mismatches 221; Indels 30; Gaps 14;

QY 1 GCCACCGCAAGCTCGCAACGCGGACACCATCACCGTCTCAACGCCATCATCAACGA 60
Db 1 GCCACTGCGACGCTCGCAACGCGGACACCATCACCGTCTCAACGCCATCATCAACGA 60
QY 61 GCGTTCCTCGGATTCCTTTTCCGAGCGCGGCGGCGGCAACCTCCGCTTCAAGACCC 120
Db 61 GCGGTTCTCGGTATTCCTTTTCTCAGCGCGGCGGCGGCAACCTCCGCTTCAAGCGCGC 120
QY 121 TGTGCGTACTCTGGCTCGCTCAACGCGCAGAGTT--ACTTACGCGCGCGTGCATGCAG 177
Db 121 TGTGCGTACTCGCGCTCTCTAATGGTCAGAAATTTACTGTATGCGCTGTGCATGCAG 180
QY 178 CAGAACCCCGAGGCGCAGTTTGAAGAGAACCTTGGCAAGACGCACTCGACTTGTGTGATG 237
Db 181 ATGAACCCATTTGGCAACTGGGACTCCTCGCTTCCCAAGGCTGCCATCAACT---TGATG 237
QY 238 CAGTCAAGGTGTTCCAGCGGTGCTTCCCGAGAGTGAAGTGCCTCAACCATCAACAGTG 297
Db 238 CAGTCCAAAGTCTTCCAGCGGTGCTTCTTCAAGCGGAGGAGTGTCTCAACCATCAACAGTG 297
QY 298 GTGCGCGCGCGGACCAAGCGGCGGCGCAACCTCCCGGTATGCTCTGGATCTTTGGC 357
Db 298 GTGCG---GCCGGGACCAAGCGGCGGCGCAACCTCCCGGTATGCTCTGGATCTTTGGC 354
QY 358 GTGGGTTTGAATCGGCGCGCGCGCAACCTTCCCTCCCGCGCGAGATGGTCAACCAAGAGT 417
Db 355 GCGGGGTTGAGTTGCGGCTCCAGTCTCTTCCCTCCCGCGCAGATGATCACCGCGCAGC 414
QY 418 GTGCTCATGGCAAG-CGATCATCCAGTGGCGGTCAATACCGTGTTCCTCGTGGGG 476
|||||
|||||
```

```
Db 415 GTGCTTATGGGCAAGCCCATCATCCAGTGAGCATGAACATACCGCGTTGCTTCGTGGGG 474
QY 477 TTCTTGGCTGTGATGACATCAAGCCGAGGCGAGCGGAAAGCCCGCTTGAAGAGACCAG 536
Db 475 TTCTTGGCTGTGTCAGACATCAAGCCGAGGCGAGCGGAAAGCCCGCTTGAAGAGACCAG 534
QY 537 CGTTTGGGCATCGAGTGGGTGGCAGACAAATTTGCGGGGTTTCGGGGCGCAACCGAGCAAG 596
Db 535 CGCTTGGGTTTGCAGTGGGTGGCGGACAAATTTGCGGGGTTTCGGGGCGCAACCGTCCAAG 594
QY 597 GTGA-CATCTTTGGGGA-GGCGGGCAGCATGTCGCTGTTGTGTCACCTCATCTGGAACGA 654
Db 595 GTGACCATCTTTGGTGAGGGCGGCGAGCATGCGGTAATGTGTACGCTCTCTGGAACGA 654
QY 655 CGGCGACAAACGTTACAAAGGCAAGCCGTTGTTTTCGCGCGGGCATCATGCGAGGAGCCAT 714
Db 655 CGGCGCAACACGTTACAAAGGCAAGCCGTTGTTTTCGTCGCGCATCATGCGAGGAGCCAT 714
QY 715 GGTGCC-GGACCCGTTGGACGGCAAGTACGGCAACGAGATCTACGACCTCTTTGTCTCGA 773
Db 715 GGTGCCGGGACCCGTTGGATGGGCCCTACGGCAAGCATCTACGACCAAGGTGTTGCTT 774
QY 774 GTGCTGGCTGTGGCAGCGCCAGGCAAGCTCGCGTCTTGGCGAGTG-CGAGCGACACC 832
Db 775 CAGCCGGCTGTGGCAGTGGCGAGCAAGCTCGCGTCTTGGCGAGCATCGAAACAA 834
QY 833 TTGCTCGATGCCACCAACAACTCTCTGGGTTCTTGGCGTACTCTCTGTTGGGTTG--- 889
Db 835 CTCCTCCAGGCCACAGCGACACTCGGGGGCTTGGCGTACCCCTCGTTGGTGGTT 894
QY 890 -TACTCCCGGCCGCGCAAGAAATCATCACCGATGACATGTAACAAGTTGGTGGCGACGG 948
Db 895 TCTCCCGCGGCCGCGCAGCACCTTCAACCATGACATGTTCAAGTTGGTGGCGACGG 954
QY 949 CAAATATGCAACGCTTCCGTCATATTGGCGACAGACAGAGGCGACCATCTTTG- 1007
Db 955 CAAATGTGCCAACGTTTCCGGTGTATTGGCGACAGAAACGAGGGGACAGTGTTCG 1014
QY 1008 GCTCTGAACGTGACCAACGATGCTCAGGCCCGTCTTACTTCAAGCAGTTTCATCCACGC 1067
Db 1015 GTTGTGAACGTGACTACGATGCTCAGGACGCGCAGTACTTCAAGGATTTATCCACGC 1074
QY 1068 CAGCGACGCGGAGATCGACACCTTGTATGGCGCGTACCCCGAGGACATCACCCAGGGTCC 1127
Db 1075 CAGCGACGCGGAGATCGACACCTTGTATGGCGCGTACCCCGAGGACATCACCCAGGGTCC 1134
QY 1128 GTTCGACACGG-----TTCAACGC--TCACCCCGCAGTTTCAAGAGAT---CGCGTGT 1177
Db 1135 GTTCGACACCGGCATCTTCAACGCCATCACCCCGCAGTTTCAACCGGATTCAGCGGTGT 1194
QY 1178 CGGCGACCTTGCATTCATCCAGCGCGCGCTACTTCTTCAACCACTTCCAGGGCGGCAC 1237
Db 1195 TGGTGAACCTTGGTTCACTCTCCCGCGGCTACTTCTCAACCACTTCCAGGGCGGCAC 1254
QY 1238 CAAATGACTGTTTCTC-AAGCAGCTCGGTTGCCAATCATGGGCACTTTCATGCAACG 1296
Db 1255 CAAATGACTGTTCTCTGAAGCAGCTGGGTTGCCCGTGTATTGGCACCCACCGCAACG 1314
QY 1297 ACATTGTGGCAGACTACTTGTGGGAAGCGGCGAGCGTCACTTACAAACACCGTTTA 1356
Db 1315 ACATTGTGGCAGACTTTTGGTGGAGCAGCAGCGCGTGTACAAACACCGTTTA 1374
QY 1357 TCGCGTTGCCACCGACTTGGACCCCAACACCGCGGGGTTGTGTGAACCTGCGCCCAAGT 1416
Db 1375 TTGCTTTTGCCAAACGACTCGAACCCGAAACAGGCGGTTTGTGTGAACCTGCGCCCAAGT 1434
QY 1417 ACACGACGCGCAGGCGCAACAACTTGTATGATCAACGCTTGGGCTTTGTACACCGGCA 1476
Db 1435 ACACGACGCGCAGGCGCAACAACTTGTGTGAGATCAACGCTTGGGCTTTGTACACCGGCA 1494
QY 1477 AGGACAACTTCCGACCCGCTGCTAGCGGCTGTGATGACCAACCGCTTCTT 1528
Db 1495 AGGACAACTTCCGACCCGCTGCTAGCGGCTGTGTTTACCAACCGCTTTT 1546
|||||
|||||
```



```

RESULT 11
CCLIPS
LOCUS
DEFINITION C.cylindracea LIP5 gene.
ACCESSION X66008.855942
VERSION X66008.1 GI:296937
KEYWORDS lip5 gene; lipase.
SOURCE Candida cylindracea
ORGANISM Candida cylindracea
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 1812)
Lotti,M., Grandori,R., Fusetti,F., Longhi,S., Brocca,S.,
Tramontano,A. and Alberghina,L.
Cloning and analysis of Candida cylindracea lipase sequences
Gene 124 (1), 45-55 (1993)
93178975
PUBMED 8440480
REFERENCE 2 (bases 1 to 1812)
Longhi,S.
Direct Submission
Submitted (06-MAY-1992) S. Longhi, Universita' degli Studi di
Milano, Via Celoria 26, 20133 Milano, ITALY
See X66006-8, X64712, X64703 & X64704
X66006, X66007 and X66008 are related sequences in the description
of C.cylindracea Lip1 and Lip2 sequences (x64703, x64704).
FEATURES
Source
Location/Qualifiers
1..1812
/organism="Candida cylindracea"
/mol_type="genomic DNA"
/strain="ATCC 14830"
/db_xref="taxon:44322"
/clone_lib="SACI genomic"
116..1765
/gene="LIP5"
116..1765
/gene="LIP5"
/EC_number="3.1.1.3"
/codon_start=1
/product="triacylglycerol lipase"
/protein_id="CAA46807.1"
/db_xref="GI:1325991"
/db_xref="GOA:P32949"
/db_xref="SWISS-PROT:P32949"
/translation="MKLALALLIASVAAPATATLANGDITGLNALINEAFILGIPFA
EPPVGNLRFKDPVYRGLNSQSFAYGPLCMQONPGTYEENLPKVALDLVMOSKVF
QAVLPNSEDCLITINVRPPGPKAGANLPVLMWIFGGFEIGSPTIFFPAQMVSKSLM
GKPIHVAVNRYLASFGFLAGPDIKAEGSSNAGLKQDLGMQWADNIAFGVDFESKV
TFGLAGSSVLCHLLWNGDNTYKGPLPRAGIMOLGAMVPLDPVDVGTGDIYDT
LVALTGSSASNKLACILGLLTQALLDNTDTPGFLSVTSRLVLPDPDGNITDM
YKVRDGRASVPIIGDNDGFLGLLLNTTTEADAEAYLRKLFPHATDADITAL
KAAVPSDVTGGLPFDPTGLNALTPQLKRINAVLGLDLPVLSRRYFLNHYTGGPKYSEL
LKQLGLPILGTHANDIVMWHFLNGSIVYNNAFAPATDLPNTAGLLVQWFKEST
SSLQAGDNLQIISALGLYTGKDNFTAGYNALFADPLHFFV"
116..250
/sig_peptide
116..250
/gene="LIP5"
251..1762
/gene="LIP5"
/product="triacylglycerol lipase"
/EC_number="3.1.1.3"
mat_peptide

ORIGIN
Query Match 64.1%; Score 982.4; DB 8; Length 1812;
Best Local Similarity 83.1%; Pred. No. 2.5e-152;
Matches 1318; Conservative 0; Mismatches 206; Indels 62; Gaps 15;

Qy 2 CCCACCGCGCAAGCTCGCCAAACGGCGACACCATCACCGGTCTCAACGCCATCATCAACGAG 61
Db 164 CCCACCGCGCGCTCGCCAAACGGCGACACCATCACCGGTCTCAACGCCATCATCAACGAG 223
Qy 62 GCGTTCTCGGCATTCCTTTTCCGAGCGCGCGGTGGGCAACCTCCGCTTCAAGACCTT 121

```

```

Db 224 GCGTTCTCGGCATTCCTTTTCCGAGCGCGCGGTGGGCAACCTCCGCTTCAAGACCTT 283
Qy 122 GTCCGCTACTCTCGCTCGCTCAACGGCCAGAGTTACTTACG-----GCCCTGTCATG 174
Db 284 GTCCGCTACCGTGGGTCTCTCAACGGTCAATCTTACCGGTACCGTCCGCTGTCGATG 343
Qy 175 CAGCAGAAACCCGAGGGCAGCTTTGAAGAGAAACCTTGGCAAGACGGCACTCGACTTGGT 234
Db 344 CAGCAGAAACCCGAGGGCAGCTTACGAGAGAACTTCCCAAGGTGGCGTTGACTTGGT 403
Qy 235 ATGCAGTCCAAAGTGTTCAGGGCGTCTTCCAGAGTGAAGTGAAGTGAAGTGAAGTGAAG 294
Db 404 ATGCAGTCCAAAGTGTTCAGGGCGTCTTCCAGAGTGAAGTGAAGTGAAGTGAAGTGAAG 463
Qy 295 GTGGTGGCGCGCGGGCAGCAACGAGGGCGCGCAACCTCCCGGTCTAGTCTCTGATCTTT 354
Db 464 GTGGTGGCGCGCGGGCAGCAACGAGGGCGCGCAACCTCCCGGTCTAGTCTCTGATCTTT 523
Qy 355 GGGCGTGGGTTCAGATCGGCGAGCCCGCACCATCTTCCCTCCCGCCAGATGGTCAACAG 414
Db 524 GCGGTGGGTTCAGATCGGCGAGCCCGCACCATCTTCCCTCCCGCTCAGATGGTCTCCAG 583
Qy 415 AGTGTGCTCATGGGCAAG-CATCATCCAGTGGCGCTCAACTACCGTGTGCTCTGTTG 473
Db 584 AGTGTGCTCATGGGCAAGCCCATCATCCAGTGGCGCTCAACTACCGTGTGCTCTGTTG 643
Qy 474 GGGTTCCTGCTGTGATGACNTCAAGCGCGAGGGCGCGGAGACCGCGGTTCGAAGGAC 533
Db 644 GGTTCCTGCGCGTCCGGACATCAAGCGCGAGGGCGAGTCCATGCGGGCTCAGAGAC 703
Qy 534 CAGCGTTTGGGCATGCAAGTGGTGGCAGACAAATTCGCGGGTTCGCGGGCGACCCGAGC 593
Db 704 CAGCGCTTGGGCATGCAAGTGGTGGCAGACAAATTCGCGGGTTCGCGGGCGACCCGAGC 763
Qy 594 AAGTGA-CATCTTTGGCGA---GGCGGGCAGCATGTCCGTTGTGTCACCTCATCTGG 649
Db 764 AAGGTGACCATCTTTTGGTGAGCTGGCGGGCAGCATGTCCGTTGTGTCACCTCATCTGG 823
Qy 650 AACGACGGCGACACAGTCAAGGGCAAGCGTGTTCGCGGGCGCATCATGCA---G 706
Db 824 AATGGCGGGCAGCAACAGCTACAGGGCAAGCGGTGTTCGCGGGCGCATCATGCACTG 883
Qy 707 GGAGCCATGTGTC---GGACCCCGTGGACGGCAGTACGACCAACAGATCTACGACCTC 763
Db 884 GGAGCCATGTGTCGCGTGGACCGGCGGACCTATGGAACCCAAATCTATGACAG 943
Qy 764 TTTGCTCGAGTGTGCTGTGGCGAGCGCGCAGCAGCAAGCTCGCGTGTGTCGCGAG---T 820
Db 944 TTGGTGGCTCTGACGGGCTGCGAGCTGCCAGCAACAGCTTGGTGTGCTTGGCGGTCTT 1003
Qy 821 GCGAGCGACACCTTGTCTCGATGCCACCAACAACTCTCTGGTTCCTTGGCGTACTCTCTCG 880
Db 1004 CTGACTCAGGATGTCTCGATGCCACCAACAGCAACCTCGGTTCCTTGTGCTGACCTCG 1063
Qy 881 TTGCGGTGTGTACT-----CCCGCGCGCAGCGGCAAGAACATCAACGATGATGTACAG 934
Db 1064 TTGCGGTGTGTACTTCCCTCCGCGCGCAGCGGCGGCAACATCAACGATGATGTACAG 1123
Qy 935 TTGTTGCGCGAGCGCAAGTATGAAAGCGTTCCTCGTATCATTTGGGACAGAAACGAGAG 994
Db 1124 TTGTTACGCGAGCGCAAGTATGAAAGCGTTCCTCGTATCATTTGGGCGACAGAAACGAGAG 1183
Qy 995 GGCACCATCTTTG-----GCTCTTGAACGTGTACACCAAGATGTCTAGGCGCGGTCTTAC 1047
Db 1184 GGTCTCTTGTGTCTCTCTCTCTGTTGAACACCAACCGAGGCGGAGCGCGAGCATAC 1243
Qy 1048 TTC---AAGCAGTTTCATCCAGCGCAGCGAGATCGACACTTGTATGGCGCGTAC 1104
Db 1244 CTCAGAAAGCTGTTTCATCCAGCGCAGCGACCGCATATCACCGATATGAAGCGCGGTAC 1303
Qy 1105 CCCAGGACATCACCCAGGT---CCGTTGACACGG-----TTCAACGCTCACCCCG 1154

```

```
Db 1304 CCCAGCGATGTCACCCAGGCTCTGCGGTTCGACACCGGCAATCTCAACGCCCTTACACCC 1363
QY 1155 CAGTTCAAGAAATC---GCGGTGCTCGGGAACCTTGATTCATCAACGCCGCCGCGCTAC 1211
Db 1364 CAGTCAAGCGGATCAATGCTGTGCTTGGGAACCTCACCTTTACCTCTCGCGCGCTAC 1423
QY 1212 TTCTCAACACCTTCCAGGCGCGACCAAGTACTCGTTCTC-----AAGCAGCTCGGG 1265
Db 1424 TTCTCAACACCTACACCGGCTGCTCCCAAGTACTCGTTCTCCTCTGAAGCAGCTTCTGGGC 1483
QY 1266 TTGCCATCATGGCACTTCCATGCCAAGCAATGTGTGGGAGACTACTTGTGGGA 1325
Db 1484 TTGCCATCTTCGCGAGCTTCCACGCGAAGCAATGTGTGGGAGCACTTTTGTGGGC 1543
QY 1326 AGCGGAGCGTCTATCAACACGCGGTTTATCGGCTTCGCCACCGACTTGGACCCCAAC 1385
Db 1544 AGCGGAGCGTCTATCAACACGCGTTCATCGGCTTGGCACCGACTTGGACCCCAAC 1603
QY 1386 ACCGCGGGTGTGTGTGAACCTGGCCCAAGTACACGAGAGCC-----AGGCAACAAC 1439
Db 1604 ACCGCGGGTGTGTGTGAAGTGGCCCAAGTCCACGAGAGCTGCGAGGCGGGGACAAC 1663
QY 1440 TTGATCATGATCAACCGCTTGGCTGTACACCGGCAAGCAACTTCCGACCGGTGGC 1499
Db 1664 TTGATGAGATGATGCTGTGGCTTGTATACCGGCAAGCAACTTCCGACCGCGGC 1723
QY 1500 TAGCAGCGGTTGATGACCAACCCGTT 1525
Db 1724 TACAAGCCTTGTTCGCGACCGCT 1749

RESULT 12
LOCUS CCLIP4
DEFINITION C.cylindracea LIP4 gene.
ACCESSION X66007.55939
VERSION X66007.1 Gi:296935
KEYWORDS lip4 gene; lipase.
SOURCE Candida cylindracea
ORGANISM Candida cylindracea
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 1785)
AUTHORS Lotti,M., Grandori,R., Fusetti,F., Longhi,S., Brocca,S.,
Tramontano,A. and Alberghina,L.
TITLE Cloning and analysis of Candida cylindracea lipase sequences
JOURNAL Gene 124 (1), 45-55 (1993)
MEDLINE 93178975
PUBMED 8440480
REFERENCE 2 (bases 1 to 1785)
AUTHORS Longhi,S.
TITLE Direct Submission
JOURNAL Submitted (06-MAY-1992) S. Longhi, Universita' degli Studi di
Milano, Via Celoria 26, 20133 Milano, ITALY
COMMENT See X66006-8, X16712, X64703 & X64704
X66006, X66007 and X66008 are related sequences in the description
of C.cylindracea Lip1 and Lip2 sequences (x64703, x64704).

FEATURES
source
1..1785
/organism="Candida cylindracea"
/mol_type="genomic DNA"
/strain="ATCC 14830"
/db_xref="taxon:44322"
/clone_lib="SACI genomic"
99..1748
/gene="LIP4"
99..1748
/gene="lip4"
/codon_start=1
/EC_number="3.1.1.3"
/product="triacylglycerol lipase"
/protein_id="CAA46806.1"
/db_xref="GI:1325990"

gene
CDS
99..1748
/gene="LIP4"
99..1748
/gene="lip4"
/codon_start=1
/EC_number="3.1.1.3"
/product="triacylglycerol lipase"
/protein_id="CAA46806.1"
/db_xref="GI:1325990"
```

```
/db_xref="GOA:P32948"
/db_xref="SWISS-PROT:P32948"
/translation="MKALVLSIVSVAAPTATLANGDTITGLNALINEAFILGIPPA
QPVGRLRPFPVPSASLNGKFTLYGLPCLQMNFELGNWDSLPKRAINLLMQSKLF
QAVLPNGEDCLTINVRPLGTGKGANLPVMVIFGGFVGGSSLPFPAQIMTASVLM
TIFELAGMSVMCKLRLMNDGNTYNGKPLFRAAIQLGAMVPLDVPDVGPGYQIYDQ
VVASCGSASDKLACLRLNDKLPQATSDTPGALAYPSLRLELPRPDGTFTIDDM
FKLVDPGKCANYPVLIIGDQDSEGTVPALLLVNTTDAQARQYFKELFHASDAEIDL
MAAPSDITQGLPFDGTGFNATLPQKRIAAVLGLDIAFTLPRIYFHHFGGKYSFL
LKQLPLGVPIGTHANDIIVQDFLVSHSVSAFNNAFANLDPNKGALLVNWPKYT
SLQLGNLLQINALGLYTKDNFTAGYDALFTNPLLFV"
99..233
/sig_peptide
/mat_peptide
/gene="LIP4"
234..1745
/gene="LIP4"
/product="triacylglycerol lipase"
/EC_number="3.1.1.3"
```

ORIGIN

```
Query Match 61.2%; Score 937.2; DB 8; Length 1785;
Best Local Similarity 81.7%; Pred. No. 7e-145;
Matches 1296; Conservative 0; Mismatches 228; Indels 62; Gaps 16;

QY 2 CCCACCGCCAAGCTCGCCAAACGCGCACACCATCACCGGTCTCAACGCCCATCATCAACGAG 61
Db 147 CCCACTGCCACGCTCGCCAAACGCGCACACCATCACCGGTCTCAACGCCCATCATCAACGAG 206
QY 62 GGGTTCTCGGCATTCCTTTGCGAGCGCGCGGTGGGCAACTCTCGCTTCAAGAGACCT 121
Db 207 GCGTTCTCGGTATTCCCTTTGCTCAGCGCGCGGTGGGCAACTCTCGCTTCAAGCGCCT 266
QY 122 GTGCGGTACTCTGGCTCGCTCAACGCCGAGAAGTT-----ACTTACGCCCGTGCATG 174
Db 267 GTGCGGTACTCTGGCTCTCTCAATGTGAGAAGTTTACTCTGTATGGCCCTCTGTGCATG 326
QY 175 CAGCAGAACCCCGAGGGCACGTTTGAAGAGAACCTTTGGCAAGACGGCACTCGACATTGGTG 234
Db 327 CAGATGAACCCATTGGGCAACTGGGACTCTCTCGCTTCCCAAGGCTGCCATCAACCTGTG 386
QY 235 ATGCAGTCAAGGTGTTCAGCGGTGTTCCTCCAGAGTAGAGACTGCTCCTACCATCAAC 294
Db 387 ATGCAGTCAAGCTCTTCCAGCGGTGTCTCTTAACGGGAGGACTGTCTCACCATCAAC 446
QY 295 GTGGTGGCGCGCGCGGACCAAGCGGCGGCACTCCCGGTGATGCTCTGGATCTTT 354
Db 447 GTGGTGGCGCGCGGTGGGACCAAGCGGTGGCAACTCCCGGTGATGCTGTGGATTTTT 506
QY 355 GCGGTGGGTGTTGAGATCGGACGCCCAACCATCTTCCCTCCCGCCCAAGATGTCACCAAG 414
Db 507 GCGCGGGGTTTGAGGTTGGCGGCTCCAGTCTCTTCCCTCCCGCACAGATGATCACCGCC 566
QY 415 AGTGTGCTCATGGGCAAG--CATCATCCAGTGGCGGTCACTACCGTGTGCTCGTG 473
Db 567 AGCGTGTCTATGGGCAAGCCCATCATCCAGTGAAGCATGAATACCGCGTGTGCTCGTG 626
QY 474 GGGTCTTGGTGGTGTGATGACATCAAGCGCGGAGGCGGCAAGCGCGCTTGAAGAC 533
Db 627 GGGTCTTGGTGGTGTGACATCAAGCGCGGAGGCGGCAAGCGCGTGTGACACGAC 686
QY 534 CAGCGTTTGGGCGATGCGAGTGGGTGGGACACACATTTGCCGGGTTCGGCGGCGACCGGAC 593
Db 687 CAACGCTTGGGTTTGACATGGGTGGGAGCAACATTTGGCGGTTCGGCGGCGACCGGTC 746
QY 594 AAGGTGA-CATCTTTGGCGA---GGCGGGCAGCATGTCTCGTGTGTGGCACTCATCTGG 649
Db 747 AAGGTGACCATCTTTGGTGAAGTGGCGGGCAGCATGTCTCGTAAATGTGTGAGTCTCTGG 806
QY 650 AACGAGCGGCAACACCGTACAGGCGGCAAGCGGTGTTCCGCGCGGCGCATCATGCA---G 706
Db 807 AACGAGCGGCAACACCGTACAGGCGGCAAGCGGTGTTCCGCGCGGCGCATCATGAGCTG 866
QY 707 GGAGCCATGGTGCC---GGACCGGTGGACGCGACGTACGGCAACGAGATCTTACGACCTC 763
```

```
Db 867 GGGGCAATGTCGCGTGGACCGTGGATGGGCTTACGGCAGCAGATCTACGACG 926
Qy 764 TTTGTCTCGAGTGTGCTGTGGCAGCGCCAGCAGCAAGTCTGGTGTCTGGCAGTGC- 822
Db 927 GTGCTTGTTCAGCGCGTGTGGCAGTGCAGCAGCAAGTCTGGTGTCTGGCAGCATC 986
Qy 823 --GAGCAGACCTTGTCTGATGTCGACCAACCAACACTCTGGTGTCTGGTGTCTGCG 880
Db 987 CTGAACGACAACTCTTCCAGGCCCAGCAGCAGCACTCCGGGGCTTGGCGTACCCCTCG 1046
Qy 881 TTGCGGTG-----TACTCTCCGCGCCGACCGCAGCAACATCATCCGATGACATGACAAG 934
Db 1047 TTGCGGTGCTGTCTCCGCGCGCCGACCGCAGCACTTCATCCGATGACATGTTCAAG 1106
Qy 935 TTGTCGCGCAGCGCAAGTATGCAAGGTTCCCGTGATCATTTGGGACGACGACGAG 994
Db 1107 TTGTCGCGCAGCGCAAGTATGCAAGGTTCCCGTGATCATTTGGGACGACGACGAG 1166
Qy 995 GGCACCATCTTTG-----GCTCTTGAACGTGACCAAGATGCTCAGGCCCGTGTTC 1047
Db 1167 GGCACAGTGTGCTGCTGCTGTGAACTGACTACCGATGCTCAGGACGCGCATC 1226
Qy 1048 TTCAAG---CAGTTTCATCCAGCGCAGCAGCGGAGATCGACACTTGTATGGCGCGTAC 1104
Db 1227 TTCAAGGACATGTTTCATCCAGCGCAGCAGCGGAGATCGACACTTGTATGGCGCGTAC 1286
Qy 1105 CCCAGGACATCACCGAGGT---CGGTTGACAGG-----TTCAAGC---TCACCCCG 1154
Db 1287 CCCAGGACATCACCGAGGTCTGCGGTTGACACCGGCACTTTTCACGCCATCACCCCG 1346
Qy 1155 CAGTTTCAGAGAAAT---CGCGGTGCTCGCGGACCTTGCAATTCATCCAGCGCGCGTAC 1211
Db 1347 CAGTTTCAGAGAAATGCGAGCGGTGCTGCTGACCTTGCTGCTCACTCTCCCGCGGCTAC 1406
Qy 1212 TTCCTCAACCACTTCCAGGGCGGCACCAAGTACTGTTTCCTC-----AAGCAGCTCGG 1265
Db 1407 TTCCTCAACCACTTCCAGGGCGGCACCAAGTACTGTTTCCTCCTTGAAGCAGCTTCTGGG 1466
Qy 1266 TTGCAATCATGGGACCTTCATGTCGACGACGACGATGTTGGCAGGACTACTTGTGGGA 1325
Db 1467 TTGCGGTGATGTCGACCCACCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1526
Qy 1326 AGCGGACGTCATCAACAGCGCTTTCGCGTTCGCGACCGACTTGCACCCCAAC 1385
Db 1527 CACAGAGCGCGGTGTAACAAACCGGTTTATGCTTTCGCAAGCAGCTGACCCGAC 1586
Qy 1386 ACCGCGGCTTGTGTTGTAACCTGCGCCAGTACACGAGCAGC-----CAGGGCAACAC 1439
Db 1587 AAGCGCGTGTGCTTGTGTAACCTGCGCCAGTACACGAGCAGCCTGCAGCTGCGCAACAC 1646
Qy 1440 TTGATGATGATCAACGCTTGGCTTGTACCGGAGGACGACCTTGGCAGCAGCTGGC 1499
Db 1647 TTGTTGAGATCAACGCTTGGCTTGTACCGGAGGACGACCTTGGCAGCAGCTGGC 1706
Qy 1500 TACGAGCGTTGATGACCAACCGCTT 1525
Db 1707 TACGAGCGTTGTTTACCAACCGCT 1732
```

```
RESULT 13
CCLIP2 2043 bp DNA linear PLN 27-APR-1993
LOCUS C.cylindracea LIP2 gene for lipase.
DEFINITION X64704
ACCESSION X64704.1 GI:2546
VERSION LIP2 gene.
KEYWORDS Candida cylindracea
SOURCE Candida cylindracea
ORGANISM Candida cylindracea
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 2043)
AUTHORS Longhi,S., Fusetti,F., Grandori,R., Lotti,M., Vanoni,M. and
```

```
Alberghina,L.
Cloning and nucleotide sequences of two lipase genes from Candida
cylindracea
Biochim. Biophys. Acta 1131 (2), 227-232 (1992)
JOURNAL 92305068
MEDLINE 1610906
PUBMED 1610906
REFERENCE 2 (bases 1 to 2043)
AUTHORS Longhi,S.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-1992) S. Longhi, Universita' degli Studi di
Milano, Via Celoria 26, 20133 Milano, ITALY
COMMENT See also X64703
X66006, X66007 and X66008 are related sequences in the description
of C.cylindracea Lip1 and Lip2 sequences (x64703, x64704).
FEATURES
source
1..2043
/organism="Candida cylindracea"
/mol_type="genomic DNA"
/strain="ATCC 14830"
/db_xref="taxon:44322"
/clone="B"
126..132
161..1807
/gene="LIP2"
161..1807
/gene="LIP2"
161..1807
/EC_number="3.1.1.3"
/cofactor="1"
/product="triacylglycerol lipase"
/protein_id="CAA45958.1"
/db_xref="GI:1325988"
/db_xref="GOA:P32946"
/db_xref="SWISS-PROT:P32946"
/translation="MKLLALGAFAAAPTATLANGDTITGLNAIVNEKPLGIPFAE
PVCTLRFKPPVPSASINGQOFTLGPLCMOMPGSFEDTLPKNARHLVLOSKIFQ
VLENDDELITINIRPPGTRASAGLPLVLMWIFGGFELGSSLPFGDMVAKSVLMG
KPVTHSVNTRVSWGFIAGEDIQNEGSHAGLDQRLAMQWADNITAGFGDPSKVT
IYGLASGASDFTVHLVWNGDNTYNGKPLFRAALQGLWVPLDFPDGTGTETINQV
VASAGCSADSKLACLRLQDLYQATSDTPGLVPLSLLLPLPDPGFTIYDDMY
ALVRDQKVAHVPIIGQNDGDTLGLLLNLTDAQARAYFKGLFHASDAEIDTLM
AAVSDTITQGLPDTGIFNATTPQKRIALLGLDPLTARRYFLNYYQGGTKYFLL
KQLGLPLVLTGHNDLIWQDYLVGSGSVLYNNAPIAFANDLDPKAGLWNTWPTYS
SLQGNLMLQINGLGLYTGKDNFEDAIYSALFSPNPLFFV"
161..202
/sig_peptide
161..202
/gene="LIP2"
/feature="triacylglycerol lipase"
203..1804
/mat_peptide
203..1804
/gene="LIP2"
/product="triacylglycerol lipase"
/EC_number="3.1.1.3"
```

ORIGIN

```
Query Match 56.6%; Score 867.6; DB 8; Length 2043;
Best Local Similarity 77.8%; Pred. No. 2.1e-133;
Matches 1233; Conservative 0; Mismatches 289; Indels 62; Gaps 13;
Qy 2 CCCACCGCCAAAGCTCGCCAAACGGGACACCATCATCCGCTCTCAACGCCATCATCAACGAG 61
Db 206 CCCACGGCCACCTCGCCAAACGGGACACCATCATCCGCTCTCAACGCCATTTGTCAACGAA 265
Qy 62 GCGTTCTCTCGGATTCCTTTGCCGAGCGCGCGGTGGGCAACCTCGCTTCAAGGACCT 121
Db 266 AAGTTCTCTCGGATACCTGTTGCCGAGCGCGCGGTGGGCAACCTCGCTTCAAGCGGCC 325
Qy 122 GTGCGGTACTCTGGCTCGCTCAACCGGCCAGAGTT-----ACTTACGGCCCGCTGCATG 174
Db 326 GTGCGGTACTCTGGCTCGCTCAACCGGCCAGAGTTTACCTGTACGGCCGCTGTGTGATG 385
Qy 175 CAGCAGAACCCCGAGGCGACGTTTCAAGAGAACCTTGGCAGACGCGACTCGACTGTGTG 234
Db 386 CAGATGAACCCCTATGGGCTCGTTTGAAGGACACACTTCCCAAGAAATCGCGGCAATTTGGTG 445
Qy 235 ATGCAGTCCAAAGGTGTTCCAGCGCGGTGTTCCCGCAGAGTGAGGAGTGCCTCACCATCAAC 294
```


Qy	1493	CGCTGGCTACGACGCGTTGATGACCAACCCGCTCTTTGTG	1532
Db	1428	GGATGCGTACAGCGCCCTCTTTTCCAAACCGCCTCTTTG	1467
RESULT 15			
AF044078		1688 bp DNA linear	SYN 07-FEB-1998
LOCUS			
DEFINITION		Synthetic construct triacylglycerol hydrolase (lip1) gene, complete	
ACCESSION	AF044078		
VERSION	AF044078.1	GI:2852389	
KEYWORDS			
SOURCE		synthetic construct	
ORGANISM		synthetic construct	
REFERENCE		artificial sequences.	
AUTHORS		1 (bases 1 to 1688)	
TITLE		Brocca, S., Schmidt-Dannert, C., Lotti, M., Alberghina, L. and Schmidt, R.D.	
JOURNAL		Design, total synthesis and functional overexpression of the	
REFERENCE		Candida rugosa lip1 gene coding for a major industrial lipase	
AUTHORS		Protein Sci. (1998) In press	
TITLE		2 (bases 1 to 1688)	
JOURNAL		Brocca, S., Schmidt-Dannert, C., Lotti, M., Alberghina, L. and Schmidt, R.D.	
REFERENCE		Direct Submission	
AUTHORS		Submitted (22-JAN-1998) Institute of Technical Biochem., University	
TITLE		of Stuttgart, Allmandring 31, Stuttgart 70569, Germany	
JOURNAL		Location/Qualifiers	
FEATURES			
source		1..1688	
		/organism="synthetic construct"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:32630"	
gene		/note="designed gene encoding lip1 of Candida rugosa"	
CDS		19..1688	
		/gene="lip1"	
		/EC_number="3.1.1.3"	
		/function="lipase"	
		/note="Lip1; codon optimized for expression in	
		Saccharomyces cerevisiae"	
		/codon_start=1	
		/product="triacylglycerol hydrolase"	
		/protein_id="AAC02093.1"	
		/db_xref="GI:2852390"	
		/translation="MELALSLIASVAAPATLANGDTITGLNAILNEAPLGIPIPA	
		EPVGNLRFDPVPSGLDQKFTSYGSCMQNPGTEYENLPKALDLVMSKVF	
		EAVSPSEDCLTINVVRPPGKAGANLPVLMWIFGGFEVGTSTFPQAMITSIAM	
		GKPIIHVSVYRVSSWGLAGDEIKAGSANAGLKQRLGMQWADNIAAFGGDPTKV	
		TIPGESAGSMVMCHILWNDDNTYKGPLFRAGIMOSGAMVSDVDGIYGNIEFDL	
		LASNAGCGSADKLALRGVSSDTLEDATNNTPGFLAYSLRLSYLPRPDGVNIITDM	
		VALVREGKYANIPIVIGDONDEGTFTGTSILNVTDAQAREVKOSFVHSDAEIDTL	
		MTATPGDITQSPFDTGILNALTQFKRISAVLDGLFTLARRVFLNHYTGGTKYSEL	
		SKQUSGLPVTGTHSDIVFDYLLGSGSLIYNNAFAITAFADLDPDNTAGLLVKKMEYIT	
		SSSQSGNNLMINALGLYTGKDNFRTAGYDALFNSNPPFFV"	
ORIGIN			
Query Match		42.5%; Score 651.8; DB 12; Length 1688;	
Best Local Similarity		69.7%; Pred. No. 8.2e-98;	
Matches 1104; Conservative		0; Mismatches 41; Indels 62; Gaps 14;	
Qy	2	CCCACCGCAAGCTCGCCACGCGGACACCATCATCCGGTCTCAACGCCATCATCAACGAG	61
Db	67	CCACCGCCACTTTGGCTAACGGTGACACCATCATCCGGTTTGAACGCCATCATCAACGA	126
Qy	62	GGCTTCTCGGATTCCTTTGGCGAGCCCGCGGTGGCAACTCCGCTTCAAGACCTT	121
Db	127	GCCTTCTGGGTATTCCATTTCGCCCAACACCACTGGTGTGATTTAGATTCAAGACCCA	186
Qy	122	GTGCGGTACTCTGGCTCGCTCAACGGCCAGAGTTACTTACGGCC-----CGTGCATG	174
Db	187	GTTCCTACTCCGGTTCCTTGGATGGTCAAAAGTTCACTTCTAGGGTCCATCTTGTATG	246

Qy 175 CAGCAGAACCCGAGGSCAGTTTGAAGAGAACCTTGGCAAGACGGCACTCGACTTGGTG 234
 Db |||||
 Qy 247 CAACAAACCCAGAGGTCCTACGAAGAAATTCGCAAGGCGAGCTTTAGATCTGGTT 306
 Db |||||
 Qy 235 ATGCAGTCCAGAGTGTTCAGGCGGTGCTCCCGAGAGTGAAGACTGCGCTCACCATCAAC 294
 Db |||||
 Qy 307 ATGCAATCCAAAGTTTTCGAAGCTGTTCCTCAATCTTCTGAAGACTGTTTGACCATTAAT 366
 Db |||||
 Qy 295 GTGGTGGCGCGCGGCGCACCAAGGGCGGCGCCAACTCCCGGTCAATGCTCTGATCTTT 354
 Db |||||
 Qy 367 GTTGTAGACACCCGCGGCAAGGCTGTGCGCAACTTGCAGGTATGTGTGATCTTT 426
 Db |||||
 Qy 355 GCGGTTGGGTTGAGATCGGAGCCGCCCAACCATCTTCCTCCCGCCAGATGGTCACCAAG 414
 Db |||||
 Qy 427 GGTGGTGGTTTGAAGTTGGTGTACTAGTAGTACCTTCCTCCAGCCCAATGATACCAAG 486
 Db |||||
 Qy 415 AGTGTGCTATGGCAGCC-ATCATCCAGTGGCCGTCAACTACCGTGTGCTCGTGG 473
 Db |||||
 Qy 487 TCTATTGCTATGGGTAAGCCCAATCATCCAGCTTCTGTCAACTACAGAGTCTCGAGCTGG 546
 Db |||||
 Qy 474 GGGTCTTGTGGCTGGTATGATGACATCAAGGCCGAGGCGAGCGGGAACGCCGCTTGAAGGAC 533
 Db |||||
 Qy 547 GGTTCCTTGGCTGGTACGAAATCAAGCCGAGAGTTCTGCCAACGCCGTTTGAAGGAC 606
 Db |||||
 Qy 534 CAGGTTTGGCARTGCAAGTGGGTGGCAGACAACATTCGCGGTTTCGCGCGGACCCGAGC 593
 Db |||||
 Qy 607 CAAAGATTGGGTATGCAATGGTGGTGGTGACAAACATTCGCTTTTGGTGGTATCCAACT 666
 Db |||||
 Qy 594 AAGGTGAC-ATCTTTGGCGA---GGCGGGAGCATGTCGCTGTGTGCGCACCTCATCTGG 649
 Db |||||
 Qy 667 AAGGTACTATCTTGGTGAATCTGCTGGTTCTATGTCCGTCATGTGTACATTTTGTGG 726
 Db |||||
 Qy 650 AACGAGCGGACAAACAGTACAAGGCGAAGCCGTTGTCGCGCGGSCATCATGCA---G 706
 Db |||||
 Qy 727 AACGAGGTGACAACTTACAGGGTACGCAATGTTTACAGCTGGTATCATGCAATCT 786
 Db |||||
 Qy 707 GGAGCCATGTGC---CGGACCGGTTGGAGCGGACGCTACGCGCAACGAGATCTACGACCTC 763
 Db |||||
 Qy 787 GGTGCTATGGTTCATCTGACGCGCTCGACGGTATCTACGGTAACGAAATTTTGTACTTG 846
 Db |||||
 Qy 764 TTTGTCTCAGTGTGCTGTGCGGCGGCGGAGGCAAGCTCGGTGCTTGGCGAGTG-- 821
 Db |||||
 Qy 847 TTGGCTTCCAAACGCTGGTGTGGTCTGCGCTCTGACAAAGTTGGCTTGTGAGAGGTGT 906
 Db |||||
 Qy 822 -CGAGCGACACCTTGTCTGATGCCAACAACTCTCGGTTCTTGGGCTACTCTCG 880
 Db |||||
 Qy 907 TCTTCTGACACTTTGGAAGAGCGCACCAACACACCCCTGTTCTTGGCTTACTCTCTCC 966
 Db |||||
 Qy 881 TTGCGGTTGTACT-----CCGCGCGCGAGCGGAAGAACATCACCGATGACATGTACAG 934
 Db |||||
 Qy 967 TTAAGATTGTCTTACTTTGCCAAGACCAGAGGTTTAAACATCACCGACGACATGTACGCT 1026
 Db |||||
 Qy 935 TTGGTGGCGAGCGCAAGTATGAAGGTTCCGCTGATCATTTGGCGACCGAAGCGAGGAG 994
 Db |||||
 Qy 1027 TTGGTTAGAGAGGTAAAGTATGCCAATCCCTGTTATCATCGGTGACCAAAACGACGAA 1086
 Db |||||
 Qy 995 GGCACCATCTTTGGCTC-----TTGAAGTGCACCAAGTGTCTCAGGCCGCTGTAC 1047
 Db |||||
 Qy 1087 GGTACCTCTTTGGTACTTCTTTGAAGCTTACCACTGATGCCAAGCCAGAGAAATAT 1146
 Db |||||
 Qy 1048 TTCAAGCA---GTTCAATCCAGCGAGCGGAGATCGACACCTTGTATGGCGGGTAC 1104
 Db |||||
 Qy 1147 TTCAAGCAATCTTTGTCCACGCTAGCGAGCTGAAATCGACACTTTGATGACTGCTTAC 1206
 Db |||||
 Qy 1105 CCCCAGACATCACCCAGGT-----CGTTTCGACACGGTTCAACGCTCACCCCG 1154
 Db |||||
 Qy 1207 CCAGGTGACATCACTCAAGGTTCTCCATTTGACACTGGAATCTAAACGCGCTTGACCCCA 1266
 Db |||||
 Qy 1155 CAGTTCAAGAGATC---GGGTGCTCGGAGCTTGCATTCATCCACGCCGCCGCTAC 1211
 Db |||||
 Qy 1267 CAATTCAAGAGAACTCTGCTGTTTGGGTGACTTGGGTTTACTTTGGCTCTGATATAC 1326
 Db |||||

Search completed: August 5, 2004, 17:10:02
 Job time : 4058 secs

Qy 1212 TTCTCAACCACTTCCAGGGCGCACCAAGTACTCGTTCTCAAGCAGCTC-----GGG 1265
 Db |||||
 Qy 1327 TTCTTGAACCACTACACCGGTGGTACCAAGTACTCTTTCTTGTCTAAGCAATTTCTTGT 1386
 Db |||||
 Qy 1266 TTGCAATCATGGGCACTTTCATGCGCAACGACATTTGTGTGGCAGGACTACTTGTGGGA 1325
 Db |||||
 Qy 1387 TTGCCAGTTTGGGTACTTTCCACTCCACGATATCGTCTTCCAAAGACTACTTGTGGGT 1446
 Db |||||
 Qy 1326 AGCGGAGCGTCTATCTACAAACGCGTTTATCGGTTCCGCCAGCTTTGGACCCCAAC 1385
 Db |||||
 Qy 1447 TCTGGTTCTTGTATCTACAAACGCTTTCATTGCTTTTGGCACTGACTTGGACCCCAAC 1506
 Db |||||
 Qy 1386 ACCGCGGGTGTGTGGTGAATCGGCCCAAGTACACC-----AGCAGCCAGGGCAACAAC 1439
 Db |||||
 Qy 1507 ACCGCGGTTTGTGTGTTAAGTGGCCAGATAACACCTTCTTCTCAATCTGGTAAACAAC 1566
 Db |||||
 Qy 1440 TTGATGATGATCAACGCGCTTGGGCTTGTACACCGGCAAGGACAACTTCCGACCGCTGGC 1499
 Db |||||
 Qy 1567 TTGATGATGATCAACGCTTTGGGTTTGTACACCGGTAAGGACAACTTTCAGAACCGCGGT 1626
 Db |||||
 Qy 1500 TACGACGCGTTGATGACCAACCC 1522
 Db |||||
 Qy 1627 TACGACGCTTGTCTCCAAACCC 1649
 Db |||||

B/gnK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 2004, 14:49:49 ; Search time 466 Seconds
(without alignments)

13966.179 Million cell updates/sec

Title: US-09-943-857-3

Perfect score: 1532

Sequence: 1 gccaccgcgaagctcgca.....tgaccacccgtttttgtg 1532

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1268	82.8	1641	7 ABX95906	Abx95906 Candida r
2	1020.4	66.6	1650	2 AAT10422	Aat10422 Candida c
3	1018.4	66.5	1641	7 ABX95909	Abx95909 Candida r
4	991.6	64.7	1649	2 AAX31111	Aax31111 Candida r
5	968.6	63.2	1641	7 ABX95908	Abx95908 Candida r
6	932	60.8	1641	7 ABX95907	Abx95907 Candida r
7	873.2	57.0	1641	7 ABX95905	Abx95905 Candida r
8	651.8	42.5	1647	2 AAX33112	Aax33112 Synthetic
9	651.8	42.5	1950	4 AAH43625	Aah43625 Synthetic
10	173	11.3	2045	2 AAQ98578	Aaq98578 Aspergill
11	173	11.3	2045	2 AAT11299	Aat11299 Aspergill
12	171.2	11.2	1674	2 AAQ05605	Aaq05605 Gene enco
13	162	10.6	1828	2 AAQ54020	Aaq54020 Lipase co
14	160.2	10.5	1635	2 AAT10421	Aat10421 Geotrichu
15	157.8	10.3	1692	2 AAQ10313	Aaq10313 Sequence
16	88.2	5.8	1716	4 AAH26618	Aah26618 Rhodospo
17	88.2	5.8	1738	2 AAQ21460	Aaq21460 Rhodospo
18	87	5.7	8528	2 AAQ46249	Aaq46249 Construct
19	62.8	4.1	2428	2 AAQ83221	Aaq83221 Bile salt
20	61.2	4.0	2184	2 AAQ71874	Aaq71874 Sequence
21	61.2	4.0	2312	6 ABQ54978	Abq54978 Human ova
22	61.2	4.0	2428	2 AAQ20452	Aaq20452 Bile-salt
23	61.2	4.0	2428	2 AAT35250	Aat35250 Human bil

24	61.2	4.0	2428	2	AAZ20300	Aaz20300 Human bil
25	61.2	4.0	2428	6	ABL70012	Abi70012 Pancreas
26	61.2	4.0	2428	6	ABA04403	Abi04403 Human mil
27	61.2	4.0	2428	6	AAS18538	Aas18538 cDNA enco
28	61.2	4.0	2487	3	AAA63883	Aaa63883 cDNA enco
29	60.6	4.0	2000	7	ADA71938	Ada71938 Rice gene
30	60	3.9	1725	3	ABA97180	Aba97180 Synthetic
31	60	3.9	1752	6	AAS17549	Aas17549 Ache codi
32	60	3.9	1845	6	AAS17493	Aas17493 Human cdn
33	60	3.9	1936	7	ACC47509	Acc47509 Human ace
34	60	3.9	2156	7	ABZ23128	Abz23128 Nucleotid
35	60	3.9	2256	2	AAQ99002	Aaq99002 Human ace
36	60	3.9	5767	6	AAS17547	Aas17547 Plasmid p
37	60	3.9	12113	6	AAS17492	Aas17492 Human ace
38	60	3.9	14446	6	AAS17548	Aas17548 Plasmid p
39	59.6	3.9	1614	2	AAQ90569	Aaq90569 Recombina
40	59.6	3.9	2166	2	AAT47862	Aat47862 Human wil
41	59.6	3.9	2166	2	AAQ90579	Aaq90579 Encodes b
42	59.6	3.9	2166	2	AAT47861	Aat47861 cDNA of h
43	59.6	3.9	2344	4	AAF31110	Aaf31110 Cholesterol
44	59.6	3.9	3004	6	ABL62109	Abi62109 Colon ade
45	59.6	3.9	3018	2	AAQ14186	Aaq14186 Human mil

ALIGNMENTS

RESULT 1
ABX95906
ID ABX95906 standard; DNA; 1641 BP.
XX
AC ABX95906;
XX
DT 15-JUL-2003 (first entry)
XX
DE Candida rugosa lipase 3 DNA.
XX
KW Lipase 3; gene; ds; mutant.
XX
OS Candida rugosa.
XX
FH Key
CDS Location/Qualifiers
FT 1..1641
FT /*tag= a
FT /product= "Lipase 3"
FT /partial
FT /note= "No start or stop codon shown"

XX
PN EP1288294-A2.
XX
PD 05-MAR-2003.
XX
PF 26-APR-2002; 2002EP-00009616.
XX
PR 31-AUG-2001; 2001US-00943857.
XX
PA (SINI-) ACAD SINICA.
XX
PI Tang S, Lee G, Shaw J;
XX
DR WPI; 2003-395476/38.
XX
PS P-PSDB; ABU09071.

Isolated mutant nucleic acid encoding Candida rugosa lipase, useful for the preparation of Candida rugosa lipase for biocatalytic applications.

Claim 25; Page 5-7; 33pp; English.

The invention relates to an isolated mutant nucleic acid encoding a Candida rugosa lipase polypeptide. The DNA has a sequence having at least 80% identity to a wild-type DNA encoding Candida rugosa lipase. The invention also relates to a microorganism comprising the DNA, where the microorganism is a bacterium or yeast, preparing a mutant DNA encoding a

CC C. rugosa lipase and a chimeric C. rugosa lipase comprising a substrate
 CC interacting domain of a first C. rugosa lipase and a non-substrate
 CC interacting domain of a second C. rugosa lipase. The method is useful for
 CC preparing a mutant DNA encoding a Candida lipase. The DNA is
 CC useful in the large scale manufacture of Candida rugosa lipase which is
 CC useful for biocatalytic applications. This sequence represents DNA
 CC encoding Candida rugosa lipase 3
 XX
 SQ Sequence 1641 BP; 315 A; 525 C; 459 G; 342 T; 0 U; 0 Other;

Query Match 82.8%; Score 1268; DB 7; Length 1641;
 Best Local Similarity 94.5%; Pred. No. 1.3e-266;
 Matches 1505; Conservative 0; Mismatches 25; Indels 62; Gaps 16;

QY 2 CCCACCGCAGCTCGCCCAACGGGACACCATCAACCGGTCTCAACGGCCATCATCAACGAG 61
 DB 43 CCCACCGCAGCTCGCCCAACGGGACACCATCAACCGGTCTCAACGGCCATCATCAACGAG 102
 QY 62 GCGTTCCTCGGCATTCCTTTGCGAGCGCGCGGTGGGCAACCTCCGCTTCAAGGACCT 121
 DB 103 GCGTTCCTCGGCATTCCTTTGCGAGCGCGGTGGGCAACCTCCGCTTCAAGGACCT 162
 QY 122 GTGCGGTACTCTGGCTCGCTCAACGGCCAGAGT-----TACTTACGGCCCG---TGCA 174
 DB 163 GTGCGGTACTCTGGCTCGCTCAACGGCCAGAGTTCACCTTCTTACGGCCCGCTTGGCAT 222
 QY 175 CAGCAGAACCCCGAGGGGACGTTTGAAGAGAACCTTGGCAAGACGGCACTCGACTGGTG 234
 DB 223 CAGCAGAACCCCGAGGGGACGTTTGAAGAGAACCTTGGCAAGACGGCACTCGACTGGTG 282
 QY 235 ATGAGTCCAAAGTGTCTCAGCGGTGCTTCCAGAGTGGAGTGGCTTCAACATCAAC 294
 DB 283 ATGAGTCCAAAGTGTCTCAGCGGTGCTTCCAGAGTGGAGTGGCTTCAACATCAAC 342
 QY 295 GTGCTGCGCGCGCGCGGACCAAGGCGGCGGCAACCTCCCGGTCACTCTGGATCTTT 354
 DB 343 GTGCTGCGCGCGCGGACCAAGGCGGCGGCAACCTCCCGGTCACTCTGGATCTTT 402
 QY 355 GCGGTGGGTGAGATGGGAGCGCGGACCATCTTCCCTCCGCGGACGATGGTCAACAG 414
 DB 403 GCGGTGGGTGAGATGGGAGCGCGGACCATCTTCCCTCCGCGGACGATGGTCAACAG 462
 QY 415 AGTGTGCTCATGGGCAAGC-CATCATCAGTGGCGGCGCAACTCCCGGTCACTCTGGATCTTT 473
 DB 463 AGTGTGCTCATGGGCAAGC-CATCATCAGTGGCGGCGCAACTCCCGGTCACTCTGGATCTTT 522
 QY 474 GGGTCTTGGCTGATGATCAATCAAGGCGGAGGCGGAAACCGCGGCTTGAAGGAC 533
 DB 523 GGGTCTTGGCTGATGATCAATCAAGGCGGAGGCGGAAACCGCGGCTTGAAGGAC 582
 QY 534 CAGCGTTGGGATGGAGTGGGTGGGAGACATTCGCGGTTCGCGGCGGACCGGAGC 593
 DB 583 CAGCGTTGGGATGGAGTGGGTGGGAGACATTCGCGGTTCGCGGCGGACCGGAGC 642
 QY 594 AAGGTGAC-ATCTTTGGCGAG--GGGAGCAGATGTCGCTGTGTCGACCTCATCTGG 649
 DB 643 AAGGTGACATCTTTGGCGAGTCTGGGCGAGCATGTCGCTGTGTCGACCTCATCTGG 702
 QY 650 AACGACGGGCAACACGTACAGGCAAGCGCGTTCGCGGCGGCGATCATGAG--- 706
 DB 703 AACGACGGGCAACACGTACAGGCAAGCGCGTTCGCGGCGGCGATCATGAGTCT 762
 QY 707 GAGGACATGGTGGCG--GACCGGTGGAGCGGACGATGAGCAACGAGATCTACGACCTC 763
 DB 763 GAGGACATGGTGGCGTGTACCGCGGTGGAGCGGACGATGAGCAACGAGATCTACGACCTC 822
 QY 764 TTTGTCTCGAGTCTGGCTGTGGCAGCGCCAGGCAAGCTCCGCTGTGGCAGTGG 823
 DB 823 TTTGTCTCGAGTCTGGCTGTGGCAGCGCCAGGCAAGCTCCGCTGTGGCAGTGG 882
 QY 824 ---AGCGACACCTTGTCTGATGCGCAACCAACACTCTTGGGTACTCTCTCG 880
 DB 883 TCTAGCGACACCTTGTCTGATGCGCAACCAACACTCTTGGGTACTCTCTCG 942

QY 881 TTGCGGTGTG-----ACTCCCGCGCGGACGGCAAGAACATCAACGATGATGATCAAG 934
 DB 943 TTGCGGTGTGTTATCTCTCCCGCGCGGACGGCAAGAACATCAACGATGATGATCAAG 1002
 QY 935 TTGCGTGGGACGGCAAGTATGCAAGGTTCCCGTGTATCATTTGGCGACCAAGACGAG 994
 DB 1003 TTGCGTGGGACGGCAAGTATGCAAGGTTTCCCGTGTATCATTTGGCGACCAAGACGAG 1062
 QY 995 GGCACCATCTTTGG-----CTCTTTGAACGTGACCAAGAAATGCTCAGGCGCGGTCTAC 1047
 DB 1063 GGCACCATCTTTGGGCTCTCTTTTGAACGTGACCAAGAAATGCTCAGGCGCGGTCTAC 1122
 QY 1048 TTCAAGCAG---TTTATTCACCGCAGGACGGGAGATCGACACCTTGTATGGCGGCTAC 1104
 DB 1123 TTCAAGCAGTCTTTTATTCACCGCAGGACGGGAGATCGACACCTTGTATGGCGGCTAC 1182
 QY 1105 CCCGAGACATCAACCCAGGG---TCCGTTTCGACACCGGTT-----CAAAGCTCACCCCG 1154
 DB 1183 CCCGAGACATCAACCCAGGGTTCTCCGTTTCGACACCGGTTTCTCAACGCGCTCACCCCG 1242
 QY 1155 CAGTTCAAGAGAAATC---GCGGTGCTCGGCAACCTTGCATTCATCCACGCGCGGCTAC 1211
 DB 1243 CAGTTCAAGAGAAATCTCTCGGTGCTCGGCAACCTTGCATTCATCCACGCGCGGCTAC 1302
 QY 1212 TTCCTCAACCACTTCCAGGGCGGCAACAGTACTCTGTTCTC-----AAGCAGTCTCGG 1265
 DB 1303 TTCCTCAACCACTTCCAGGGCGGCAACAGTACTCTGTTCTCTTAAGCAGTCTCTGTTGG 1362
 QY 1266 TTGCAATCATGGGACCTTCCATGCAACGACATTTGTGGCAGGACTACTTGTGGGA 1325
 DB 1363 TTGCAATCATGGGACCTTCCATGCAACGACATTTGTGGCAGGACTACTTGTGGGA 1422
 QY 1326 AGCGGACGCTCATCTCAACCAACGCGTTTATCGGTTTCGCCACGACTTGGACCCCAAC 1385
 DB 1423 AGCGGACGCTCATCTCAACCAACGCGTTTATCGGTTTCGCCACGACTTGGACCCCAAC 1482
 QY 1386 ACCCGGGGTTTGGTGAACCTGGCCCAAGTATACACGACG-----CAGGGCAACAC 1439
 DB 1483 ACCCGGGGTTTGGTGAACCTGGCCCAAGTATACACGACG-----CAGGGCAACAC 1542
 QY 1440 TTGATGATCATGACGCTTGGGCTTGTACACCGCAAGGACAACTTCCGACCGCTGGC 1499
 DB 1543 TTGATGATCATGACGCTTGGGCTTGTACACCGCAAGGACAACTTCCGACCGCTGGC 1602
 QY 1500 TACGACGCTTGTGATGACCAACCCGTTCTTTGT 1531
 DB 1603 TACGACGCTTGTGATGACCAACCCGTTCTTTT 1634

RESULT 2

AAT10422
 ID AAT10422 standard; DNA; 1650 BP.
 XX AAT10422;
 AC AAT10422;
 XX 20-SEP-1996 (first entry)
 DT
 XX Candida cylindracea lipase gene.
 DE
 XX Fatty acid; transgenic plant; exogenous; lipase; Rhizopus niveus;
 KW Pseudomonas aeruginosa; Pseudomonas fluorescens; Geotrichum candidum;
 KW Candida cylindracea; constitutive; tissue specific; promoter; lipid;
 KW milling; biofuel; lubricant; detergent; ss.
 XX
 OS Candida cylindracea.
 XX Key Location/Qualifiers
 FH 46..1647
 FT /*tag= a
 FT /product= "lipase enzyme"
 XX
 PN FR272798-A1.

XX	26-JAN-1996.	Db	589	CAGCGCTTGGGATGACGAGTGGTGGGCAACAATTCGGCGGCTTTGGCGGACCCGACC	648
XX		Qy	594	AAGTGA-CATCTTTGGCGA---GGCGGAGAGATGTCGGTGTGGTGGCCACTCATCTGG	649
XX	25-JUL-1994; 94FR-00009272.	Db	649	AAGTGAACATCTTTGGCGGAGCTGGCGGAGAGATGTCGGTGTGGTGGCCACTCATCTGG	708
XX	25-JUL-1994; 94FR-00009272.	Qy	650	AAGCGCGGACAAACGCTAAGAGGCAAGCGTGTTCGGCGGCGGCATCATGCA---G	706
XX	(NAPO-) INST NAT POLYTECHNIQUE TOULOUSE.	Db	709	AAGCGCGGACAAACGCTAAGAGGCAAGCGTGTTCGGCGGCGGCATCATGCACTG	768
XX	Alibert G, Mouloungui Z, Boudet A;	Qy	707	GGAGCCATGTGTC---GGACCCGGTGGACGGCAAGCTACGGCAACGAGATCTACGACTC	763
XX	WPI; 1996-107680/12.	Db	769	GGGCGCATGTGTCGGTGGACGGCGTGGACGGCATCTACGGCAACGAGATCTTTGACCTC	828
XX	Prodn. of fatty acids or derivs. from transgenic oilseed plants -	Qy	764	TTTGTCTCGAGTGTGCTGGTGGACGGCGGAGGCAAGCTCGGTGCTTGGCGAGTGC-	822
XX	engineered to express a lipase that contacts lipid(s) only when seeds are	Db	829	TTGGCGTGAACGCGGGCTGCGGAGCGCGGAGCAAGCTTGGTGTGCTTGGCGGGTGTG	888
XX	milled.	Qy	823	--GAGCGACACCTTGTCTGATGCCAACCAACACTCTCTGGTCTTGGGCTACTCTCG	880
XX	Claim 9; Page 26-27; 32pp; French.	Db	889	CTGAGCGACAGCTTGGAGGAGCGCAACAACAACCCCTGGTCTTGGCGTACTCTCTCG	948
XX	A novel method of producing fatty acids or their derivs. in a plant	Qy	881	TTGCGGTG-----TACTCCCGCGCGGAGCAAGCAATCACCGATGACATGACAAG	934
XX	comprises generating a transgenic plant contg. an exogenous lipase gene,	Db	949	TTGCGGTGCTGTGTACCTCCCGCGCGGAGCAAGCAATCACCGATGACATGACAAG	1008
XX	esp. selected from the Rhizopus niveus, Pseudomonas aeruginosa,	Qy	935	TTGGTGGGAGCGGCAAGTATGCAAGGTTCCCGTGTATCTTGGCGACCAAGACGAGAG	994
XX	P.fluorescens, Pseudomonas sp., Geotrichum candidum or Candida	Db	1009	TTGGTGGGAGCGGCAAGTATGCAAGTATGCAAGTATGCAAGTATGCAAGTATGCAAG	1068
XX	cylindracea lipase genes (AAT10417-22 resp.). The lipase genes can be	Qy	995	GGCACCATCTTTG-----GCTCTTGAACGTCACCAACACTCTCTGGTCTTGGGCTACTCTCG	1047
XX	placed under control of a constitutive or a tissue specific promoter. The	Db	1069	GGCACCTTCTTTGGGACCCCTGCTGTGTGAAGTGAACGATGCCAGGCGCCGAGTAC	1128
XX	production of the fatty acids only occurs when the lipase and lipids	Qy	1048	TTCAAGCA---GTTTCAACGCGGAGGAGATGACACCTTGTATGGCGGGTAC	1104
XX	contact each other after milling of the plants. The fatty acids generated	Db	1129	TTCAAGCAGCTGTTTGTCCAGCGGAGCGGAGATGACACGTTGATGACGCGGTAC	1188
XX	can be used to prod. e.g. biofuels, lubricants, detergents, etc	Qy	1105	CCCCAGGACATCACCCAGGG---TCCGTTGACACGCGTT-----CAAGCTCACCCCG	1154
XX	Sequence 1650 BP; 299 A; 540 C; 491 G; 320 T; 0 U; 0 Other;	Db	1189	CCCGGCGCATGACCCAGGGCGTCCGCTTTCGACACGCGGTATTCTCAACGCGCTTACCCCG	1248
XX	Query Match 66.8%; Score 1020.4; DB 2; Length 1650;	Qy	1155	CAGTTCAAGAGAAATC---GCGGTGCTGCGGACCTTTCATTCATCCACGCGCGCGCTAC	1211
XX	Best Local Similarity 84.7%; Pred. No. 1.2e-212;	Db	1249	CAGTTCAAGAGAAATCCTGCGGCTGCTGCGGACCTTGGCTTTACGCTTGTCTCGCTAC	1308
XX	Matches 1341; Conservative 0; Mismatches 181; Indels 62; Gaps 15;	Qy	1212	TTCTCAACCACTTCCAGGGCGGACCAAGTACTCGTTCTCTC-----AAGCAGCTCGGG	1265
QY	2 CCACCGCAAGCTCCGACGCGGACCAACCATCACCGTCTCAACGCCATCATCAACGAG	Db	1309	TTCTCAACCACTACACCGCGGCGCAAGTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1368
Db	49 CCACCGCACGCTCCGACGCGGACCAACCATCACCGTCTCAACGCCATCATCAACGAG	Qy	1266	TTGCCAATCATGGGACCTTCCATGCGCAACGACATGTGTGGCAGGACTACTTGTGGGA	1325
QY	62 GCGTCTCTCGGATTCCTTTGCGGAGCGCGGCGGCAACCTCCGCTTCAGGACCTT	Db	1369	TTGCCGCTGCTCGGAACGTTCCACTCCAAACGACATTTGCTTTCAGGACTACTTGTGGG	1428
Db	109 GCGTCTCTCGGATTCCTTTGCGGAGCGCGGCGGCAACCTCCGCTTCAGGACCTT	Qy	1326	AGCGGCGAGCGTCTATCTACAAACAAACCGGTTATGCGGTTGCGGCGGAGCTTGGACCCCAAC	1385
QY	122 GTGCGCTACTGTGGCTCGCTCAACGCGGACCAAGTT-----ACTTACGCGCGCTGCATG	Db	1429	AGCGGCTCGCTCATCTACAAACAAACCGGTTATGCGGTTGCGGCGGAGCTTGGACCCCAAC	1488
Db	169 GTGCGCTACTGTGGCTCGCTCAACGCGGACCAAGTTCAACGCCATCATCAACGAG	Qy	1386	ACCGGGGCTTGTGTGTAACCTGGGCTTGTACCCGCGGAGGACAACTTCCGACCGCTGGC	1439
QY	175 CAGCAGACCCGAGGCGAGCTTGAAGAGACCTTGGCAAGCGCATCTGACTTGGTG	Db	1489	ACCGGGGCTTGTGTGTAACCTGGGCTTGTACCCGCGGAGGACAACTTCCGACCGCTGGC	1548
Db	229 CAGCAGACCCGAGGCGACCTTACGAGGAGAACCTCCCAAGGCGAGCTCGACTTGGTG	Qy	1440	TTGATGATGATCAACGCGCTTGGGCTTGTACCCGCGGAGGACAACTTCCGACCGCTGGC	1499
QY	235 ATGCGATCCAAGTGTTCAGGCGGTGCTTCCAGAGTGAGGACTGCTCACCATCAAC	Db	1500	TACGACGCGTTCATGACCAACCCG	1523
Db	289 ATGCGATCCAAGTGTTCAGGCGGTGCTTCCAGAGTGAGGACTGCTCACCATCAAC	Qy	1609	TACGACGCGTTCATGACCAACCCG	1632
QY	295 GTGGTGGCGCGCGGCGACCAAGCGGCGGCAACCTCCCGGTCTATGCTTGGATCTTT	Db			
Db	349 GTGGTGGCGCGCGGCGACCAAGCGGCGGTGCAACCTCCCGGTGATGCTTGGATCTTT	Qy			
QY	355 GCGGCTGGGTTGAGATCGGAGCGCCACCATCTTCCCTCCGCGCGAGATGTGTCACCAAG	Db			
Db	409 GCGGCGGCTTGTGAGTGGTGGCGACCAAGCAGCTTCCCTCCGCGCGAGATGATCAACAG	Qy			
QY	415 AGTGTGCTCATGGGCAAG-CCATCATCCACGTGGCGTCAACTTACCGTGTTCGCTCGTG	Db			
Db	469 ASCATGTCATGGGCAAGCGCCATCATCCAGTGTGAGGCTCAACTTACCGGTGCTGCTGG	Qy			
QY	474 GGGTCTTGGCTGGTATGATCAACAGCGCGGAGGCGGGAACCGCGGCTTGAAGGAC	Db			
Db	529 GGGTCTTGGCTGGGACGAGATCAAGCGGAGGCGAGTCCCAACGCGGTTGAAGGAC	Qy			
QY	534 CAGCGTTGGGATGAGTGGTGGCAGACAACTTCCGCGGTTCGCGGCGACCCGAGC	Db			

ID ABX95909 standard; DNA; 1641 BP.

AC AC

DT 15-JUL-2003 (first entry)

DE Candida rugosa lipase 8 DNA.

XX Lipase 8; gene; ds; mutant.

XX Candida rugosa.

XX Key Location/Qualifiers

FT 1.1641

FT /*tag= a

FT /product= "Lipase 8"

FT /partial

FT /note= "No start or stop codon shown"

XX EPI288294-A2.

XX 05-MAR-2003.

XX 26-APR-2002; 2002EP-00009616.

XX 31-AUG-2001; 2001US-00943857.

XX (SINI-) ACAD SINICA.

XX Tang S, Lee G, Shaw J;

XX WPI; 2003-395476/38.

XX P-PSDB; ABU09074.

XX Isolated mutant nucleic acid encoding Candida rugosa lipase, useful for

XX the preparation of Candida rugosa lipase for biocatalytic applications.

XX Claim 25; Page 10-11; 33pp; English.

XX The invention relates to an isolated mutant nucleic acid encoding a

XX Candida rugosa lipase polypeptide. The DNA has a sequence having at least

XX 80% identity to a wild-type DNA encoding Candida rugosa lipase. The

XX invention also relates to a microorganism comprising the DNA, where the

XX microorganism is a bacterium or yeast, preparing a mutant DNA encoding a

XX C.rugosa lipase and a chimeric C.rugosa lipase comprising a substrate

XX interacting domain of a first C.rugosa lipase and a non-substrate

XX preparing a mutant DNA encoding a Candida rugosa lipase. The DNA is

XX useful in the large scale manufacture of Candida rugosa lipase which is

XX useful for biocatalytic applications. This sequence represents DNA

XX encoding Candida rugosa lipase 8

XX Sequence 1641 BP; 298 A; 537 C; 472 G; 334 T; 0 U; 0 Other;

XX Query Match 66.5%; Score 1018.4; DB 7; Length 1641;

XX Best Local Similarity 84.7%; Pred. No. 3.2e-212;

XX Matches 1349; Conservative 0; Mismatches 181; Indels 62; Gaps 16;

235 ATGAGTCCAAAGGTGTTCCAGGCGGTGCTTCCAGAGAGTGGAGTGCCTCACCATCAAC 294

283 ATGAGTCCAAAGGTGTTGAGGCGGTGCTCCGTAGCGAGGACTGTCTCACCATCAAC 342

295 GTGGTGGCGCCCGCGGACCAAGGGGGGGGCAACCTCCGGTCTAGTCTGGATCTTT 354

343 GTGGTGGCGCCCGCGGACCAAGGGGGGTGCAACCTCCGGTGTGCTCTGGATCTTT 402

355 GCGGTGGGTGTTGAGATCGGACGCCACCACATCTTCCCTCCCGCCAGATGTCACCAAG 414

403 GCGGTGGGTGTTGAGGTGGTGGACCAAGCACCTTCCCTCCCGCCAGATGTCACCAAG 462

415 AGTGTGCTCATGGGCAAG-CCATCATCCAGTGGCGCTCAACTACCGTGTCTGTGG 473

463 AGCATTTGCCATGGCAAGCCCATCATCCAGTGGCGCTCAACTACCGTGTCTGTGG 522

474 GGGTTCCTGGTGTGATGATCATCAAGCCGAGGCGGAGGAGCGCGGCTTGAAGGAC 533

523 GGGTTCCTGGTGTGATGATCATCAAGCCGAGGCGGAGTGCACACGCGGCTTGAAGGAC 582

534 CAGGTTTGGGTCATGGTGGGAGCAACAATTCGCGGTTTCGGCGGCGACCCGAGC 593

583 CAGGTCATGGGTCATGGTGGGAGCAACAATTCGCGGTTTCGGCGGCGACCCGAGC 642

594 AAGGTGA-CATCTTTGGCGAG---GCGGGGAGCATGTCCGGTGTGTCGACCTCATCTGG 649

643 AAGGTGACCATCTTTGGCGAGTCTGCGGCGAGCATGTCCGGTGTGTCGACCTCATCTGG 702

650 AACGACGGGACAAACAGTACAGGGCAAGCGGTGTTCCGCGGCGCATCATCGAG--- 706

703 AACGACGGGACAAACAGTACAGGGCAAGCGGTGTTCCGCGGCGCATCATCGAGTCT 762

707 GGAGCCATGTT---GCGGACCCCGTGGACGCGGAGTACGCGGAGATCTACGACCTC 763

763 GGGGCCATGTTACCGTGGAGCGCGGTGGACGCGGTCTACGCGAGGAGATCTTTGACCTC 822

764 TTTGCTCGAGTGTGGTGGGAGCGCGGAGCAAGTCTGGTGTGTCGCGAGTCTGG 823

823 TTGGCTCGGAGCGCGGCTGGCGAGCGGCAAGTCTGGTGTGTCGCGGCTGG 882

824 ---AGGACACCTTGTCTGATGCGGACCAACCAACTCTCTGGTCTTCTGGCTACTCTCTCG 880

883 TCTAGCGACCTTGGAGCGCGGACCAACCAACTCTCTGGTCTTCTGGCTACTCTCTCG 942

881 TTGGCTGTGT---ACTCCGCGCGGAGCGGAGCAACATCATCGGATGATGATGACG 934

943 TTGGCTGTGTCTTATCTCTCCGCGCGGAGCGGCTGAACTACCGGAGATGTTTGGC 1002

935 TTGGTGGCGGACGGCAAGTATGCAAGCGTTCGCGTGTGATCTTTGGCGAGCGAGACGAG 994

1003 TTGGTCCGAGGGCAAGTATGCAAGCGTTCGCGTGTGATCTGCGGCGAGCAAGACGAG 1062

995 GGCACCATCTTTGGCTC-----TTGAACTGACCAAGATGCTCAGGCCCGTCTTAC 1047

1063 GGCACCTCTTTGGCACCTCTTCTTTGAACTGACCAAGATGCGGAGCGCGGAGCGGAGTAC 1122

1048 TTCAAGCAG---TTCATCCAGCGGAGCGGAGATCGACACCTTGTATGCGCGGCTAC 1104

1123 TTACGCGAGTCTTTTGTCCAGCGGAGCGGAGTCTGACACGTTGATGACGCGGAT 1182

1105 CCCGAGGACATCACCCAGGG---TCCGTTGGACAGCGGTT-----CAAAGTCAACCGG 1154

1183 CCCGAGGACATCACCCAGGGTTCTCCGTTTCCAGACGCGGTTTCTCAAAGCGCTCACCCCG 1242

1155 CAGTTCAAGAGATC---GCGGTCTCGCGGAGTTCATCTCATCCAGCGCGCGGCTAC 1211

1243 CAGTTCAAGAGATCTCTCGGCTGCTCGGCGAGCTTGCCTCTCATCCAGCGCGGCTGCTAC 1302

1212 TTCTCTCAACCACTTCAGGCGGCGGAGTACTCGTCTCTC-----AAGCAGTCTGGG 1265

1303 TTCTCTCAACCACTACAGCGGCGGAGTACTCATCTCTCTAAGCAGGCTCTCTGGC 1362

QY 1266 TTGCAATATGCGGACCTTCCATGCGCAACGACATTTGTGGGAGGACTACTTTGTGGGA 1325
 DB 1363 TTGCGGCTGCTCGGAAAGTTTCCATCTCCAAAGACATTTGTTCAGGACTACTTTGTGGGC 1422
 QY 1326 AGCGGACGCTCATCTACAAACGCGTTTATCGGTTCCGACCGACTTGGACCCCAAC 1395
 DB 1423 AGCGGCTGCGTCACTCTACAAACGCGTTTATCGGTTTCCACGGAATTTGGACCCCAAC 1482
 QY 1386 ACCGCGGGTTGTTGGTGAATCTGCCCAAGTACACCAAGCAGC-----CAGGGCAACAAC 1439
 DB 1483 ACCGCGGGTTGTTGGTGAAGTGGCCGAGTACACCAAGCAGCTCTCAGTCTGGCAACAAC 1542
 QY 1440 TTGATGATGATCAACGCTTGGGTTGTATACCGGCAAGACAACTTCGACCGCTGGC 1499
 DB 1543 TTGATGATGATCAACGCTTGGGTTGTATACCGGCAAGACAACTTCGACCGCTGGC 1602
 QY 1500 TACGACGCTGTGATGACCAACCGTTCTTTGT 1531
 DB 1603 TACGACGCTGTGTTCTCAACCGCCGCTCTTT 1634

RESULT 4

AAX33111

ID AAX33111 standard; DNA; 1649 BP.

XX AAX33111;

XX

23-JUN-1999 (first entry)

XX

Candida rugosa lipase 1 gene.

XX

Candida rugosa; lipase 1; LIPI; industrial bioconversion; ss.

XX

Candida rugosa.

XX

WO9914338-A1.

XX

25-MAR-1999.

XX

16-SEP-1997; 97WO-NL000524.

XX

16-SEP-1997; 97WO-NL000524.

XX

(UNIL) UNILEVER NV.

XX

Brocca S, Schmidt-Dannert C, Lotti M, Alberghina L, Schmid R;

XX

WPI; 1999-229539/19.

XX

Synthesis and functional overexpression of a Candida rugosa lipase gene coding for a major industrial lipase.

XX

Claim 13; Page 23-28; 44pp; English.

XX

CC The present sequence represents the native Candida rugosa lipase 1 gene.
 CC lipases produced by Candida rugosa are extensively used in industrial
 CC bioconversions, and the pure lipase 1 can be used in a process requiring
 CC high specificity toward acyl chains shorter than 14C. Lipase 1, free of 2
 CC -5, can be obtained without using extensive and expensive working up
 CC procedures. Pure lipase 1 exhibits higher activity toward capriate than
 CC toward palmitate

XX

SQ Sequence 1649 BP; 317 A; 536 C; 493 G; 301 T; 0 U; 2 Other;

Query Match

Best Local Similarity 64.7%; Score 991.6; DB 2; Length 1649;

Matches 1336; Conservative 0; Mismatches 186; Indels 64; Gaps 17;

QY

2 CCACCGCCCAAGCTCGCCAAACGCGGACACCATCACCGGTTCTCAACGCCATCATCAACGAG 61
 DB 49 CCACCGCCCAAGCTCGCCAAACGCGGACACCATCACCGGTTCTCAACGCCATCATCAACGAG 108

QY

62 GGGTTCTCGGCATTCCTTTGCGGAGCGCGGGTGGCAACCTCGCTTCAAGGACCCT 121

DB 109 GCGTTCCTCGGCATTCCTTTGCGAGCGCGCGTGGCAACCTCCGCTTCAAGGACCCC 168
 QY 122 GTGCGGTACTCTGCTCGCTCAACGGCCAGAAGTT-----ACTTACGGCCCG---TGCATG 174
 DB 169 GTGCGGTACTCTGCTCGCTCGATGGCCAGAAGTTTACGAGCTACGCGCCCGAGCTGATG 228
 QY 175 CAGCAGAAACCCCGAGGCGCAGCTTTTGAAGAGAACCTTGGCAAGCGGACCTCGACTCTGGTG 234
 DB 229 CAGCAGAAACCCCGAGGCGCAGCTTACGAGAGAACCTCCCAAGGCGGCTGACTTGGTG 288
 QY 235 ATGCACTCCAAGGTGTTTCCAGGCGGTCTTCCCAGAG--TGAGGACTCCCTCACCATCAA 293
 DB 289 ATGCACTCCAAGGTGTTTGAAGCGGTGAGCCGCGGAGCAGCAGGAGACTGTCTCACCATCAA 348
 QY 294 CGTGTGTCGCGCGCGCGGCAACCAAGGCGGCGGCAACCTCCCGGTCAATGTCTGGATCTT 353
 DB 349 CGTGTGTCGCGCGCGCGGCAACCAAGGCGGCTGCAACCTCCCGGTGATGTCTGGATCTT 408
 QY 354 TGGCGGTGGGTGTTGAGATCGGACGCCACCATCTTCCCTCCCGCCAGATGTCACCAA 413
 DB 409 TGGCGGTGGGTGTTGAGGTGGTGGCACGACCTTCCCTCCCGCCAGATGTCACCAA 468
 QY 414 GAGTGTCTCATGCGGCAAG--CCATCATCTCACTGCGCGTCAACTACCGTGTGCTCTGTG 472
 DB 469 GAGCATTCGATGGCAAGCCCATCTCCAGGTGAGCGTCAACTACCGTGTGCTCTGTG 528
 QY 473 GGGGTCTTGGCTGGTGTGATGACATCAAGGCGGAGGCGGCGGAAACCGCGGCTTGAAGGA 532
 DB 529 GGGGTCTTGGCTGGGCGACGAGATCAAGGCGGCGGCACTGCGCAACCGCGGCTTGAAGGA 588
 QY 533 CCAGCGTTTGGGCGATCGAGTGGGTGCGAGCAACATTTGCCGCTTCCGCGCGGCGACCGGAG 592
 DB 589 CCAGCGTTTGGGTATGCAGTGGGTGGGCAACATTTGCCGCTTCCGCGCGGCGACCGGAC 648
 QY 593 CAAAGTGCAC--ATCTTTGGCGAG---GCGGCGAGCATGTCCGCTTGTGCGCACCTCATCTG 648
 DB 649 CAAAGTGCAGATCTTTGGCGAGAGCGGCGGCGAGCATGTCCGCTCATGTGCCACATTTCTG 708
 QY 649 GAACGAGCGGCGCAACACGCTAAGGGCAAGCGGTTGTTCCGCGCGGCGCATATGAGCAG-- 706
 DB 709 GAACGAGCGGCGCAACACGCTAAGGGCAAGCGGCTTCCGCGCGGCGCATATGAGCAG 768
 QY 707 -GGAGCCATGTGCGCG---GACCCGCTGAGCGGCGAGCATGCGGCAACGAGATCTACGACCT 762
 DB 769 CGGGCGATGTGTCGCGAGCGACGCGGTGACGCGCATCTACGCAACGAGATTTTGCCT 828
 QY 763 CTTTGTCTCGAGTGTGCTGTGGCAGCGCCAGCGACAAGCTCGCGTGTGCGCAGTGC 822
 DB 829 CTTTGGCGTCGAACGCGGCGTGGCGAGCGCCAGCGACAAGCTCGCGTGTGCGCGGTGT 888
 QY 823 G---AGCGACACTTGTCTGATGCGACCAACACTCTCTGGGTTCTTGGCGTACTCCTC 879
 DB 889 GAGCAGCAGCAGCTTGGAGGAGCGCCACCAACAAACACCCCTGGGTTCTTGGCGTACTCCTC 948
 QY 880 GTTTCGGTTGTACT-----CCGCGCCCGACGCGAGACATCACCGATGACATGTACAA 933
 DB 949 GTTTCGGTTGAGCTACTCTCCCGCGCCCGCGGCGTGAACATCACCGAGCATGTACGC 1008
 QY 934 GTTGTGCGCGCAGCGGCAAGTATGCAAGCGTTTCCCGTGTATCTTGGCGCACCAAGACCA 993
 DB 1009 CTTTGTGCGCGAGGCGCAAGTATGCCAACATCTCTGTGATCATCGGCGCACCAAGACCA 1068
 QY 994 GGGCACCATCTTTGGC-----TCTTGAACGTGACCAAGATGTTCAGCCCGCTGTCTTA 1046
 DB 1069 GGGCACCTCTTTTGGCGACCGAGCAGCTTGAACGTGACCAAGATGCGCGCGCGGAGTA 1128
 QY 1047 CTTCAAGCAG---TTTATCCAGCGCAGCGAGGATGACACCTTGTATGGCGCGTA 1103
 DB 1129 TTTCAAGCAGAGCTTTGTCCACGCGGAGCGGAGATCGACACGTTGATACGCGCGTA 1188
 QY 1104 CCCCAGGACATCACCCAGG---GTCCGCTTCCACACGCTT-----CAACGCTCACCCC 1153

Db 1189 CCCGCGGATCATCACCCAGGCGACCGCGTTTCGACACGGGTATTCTCAACGGCTTCAACCC 1248
 Qy 1154 GCAGTTCAAGAGAT---CGGFGTCGCGGACCTTGCATTCTCAACCGCCGCGCTA 1210
 Db 1249 GCAGTTCAAGAGATCAGCGCGGTCTCGCGACCTTGGCTTTACGCTTGTCTGCTGCTA 1308
 Qy 1211 CTTCCTCAACACACTTCCAGGCGGCGACCAAGTACTCGTTCTCA-----AGCAGCTCG 1254
 Db 1309 CTTCCTCAACACACTTCCAGGCGGCGACCAAGTACTCGTTCTCAACGAGCTCAGCG 1368
 Qy 1265 GTTCCCAATCATGGGCGCTTCCATGCCAAGCAATGTGTGGGAGGACTACTTGTGGG 1324
 Db 1369 CTTCGCGGTCTCGGAAGCTTCCACTCCACGACATGTCTTCCAGGACTACTTGTGGG 1428
 Qy 1325 AAGCGGAGCGGTCTACTACACAGCGGTTCATCGCTTCCGACCGACTTGGACCCAA 1384
 Db 1429 CAGCGGCTCGCTCATCTACAAACAGCGGTTCATCGCTTCCGACCGACTTGGACCCAA 1488
 Qy 1385 CACCGCGGGTGTGTGTGAACCTGCGCCCAAGTACACAGCAGC-----CAGGCGCAACA 1437
 Db 1489 CACCGCGGGTGTGTGTGAAGTGGCGCGAGTACACAGCAGCAGCGCAGCGCAACA 1548
 Qy 1438 ACTTGATGATCAACAGCTTGGGCTTGTATACCGGCAAGCAACTTCCGACCGCTG 1497
 Db 1549 ACTTGATGATCAACAGCTTGGGCTTGTATACCGGCAAGCAACTTCCGACCGCTG 1608
 Qy 1498 GCTACGAGCGGTGATGACCAACCG 1523
 Db 1609 GCTACGAGCGGTGTGTTCACCAACCG 1634

RESULT 5

ABX95908
 ID ABX95908 standard; DNA; 1641 BP.

AC AC

ABX95908;

15-JUL-2003 (first entry)

Candida rugosa lipase 5 DNA.

Lipase 5; gene; ds; mutant.

Candida rugosa.

Key Location/Qualifiers

1..1641

/*tag= a

/product= "Lipase 5"

/partial

/note= "No start or stop codon shown"

EP1288294-A2.

05-MAR-2003.

26-APR-2002; 2002EP-00009616.

31-AUG-2001; 2001US-00943857.

(SINI-) ACAD SINICA.

Tang S, Lee G, Shaw J;

WPI; 2003-395476/38.

P-P8DB; ABU09073.

Isolated mutant nucleic acid encoding Candida rugosa lipase, useful for the preparation of Candida rugosa lipase for biocatalytic applications.

Claim 25; Page 8-9; 33pp; English.

The invention relates to an isolated mutant nucleic acid encoding a

CC Candida rugosa lipase polypeptide. The DNA has a sequence having at least 80% identity to a wild-type DNA encoding Candida rugosa lipase. The CC invention also relates to a microorganism comprising the DNA, where the CC microorganism is a bacterium or yeast, preparing a mutant DNA encoding a CC C. rugosa lipase and a chimeric C. rugosa lipase comprising a substrate CC interacting domain of a first C. rugosa lipase and a non-substrate CC interacting domain of a second C. rugosa lipase. The method is useful for CC preparing a mutant DNA encoding a Candida rugosa lipase. The DNA is CC useful in the large scale manufacture of Candida rugosa lipase which is CC useful for biocatalytic applications. This sequence represents DNA CC encoding Candida rugosa lipase 5

XX Sequence 1641 BP; 302 A; 543 C; 444 G; 352 T; 0 U; 0 Other;

Query Match 63.2%; Score 968.6; DB 7; Length 1641;

Best Local Similarity 82.6%; Pred. No. 2.3e-201;

Matches 1309; Conservative 0; Mismatches 214; Indels 62; Gaps 15;

Qy 2 CCCACCGCCAAAGCTCGCCAAACGGCGACACCATCAACCGTCTCAACGGCATCATCAACGAG 61

Db 43 CCCACTGCCAGCTCGCCAAACGGCGACACCATCAACCGTCTCAACGGCATCATCAACGAG 102

Qy 62 GCGTTCCTCGGCATTCCCTTTGCCAGCGCGGTGGGCAACCTCCGCTTCAAGGACCCCT 121

Db 103 GCGTTCCTCGGCATTCCCTTTGCCAGCGCGGTGGGCAACCTCCGCTTCAAGGACCCCT 162

Qy 122 GTGCCGTACTCTGGCTCGCTCAACGGCGCAAGTAATTCTTACGGCC-----CGTGCATG 174

Db 163 GTGCCGTACTCTGGCTCGCTCAACGGTCAATCTTCAACGGTCAAGGTCCTTCTTGCATG 222

Qy 175 CAGCAGAACCCGAGGGGACGTTTGAAGAGAACCTTGGCAGAGCGGCACTCGACTGTGGT 234

Db 223 CAGCAGAACCCGAGGGGACCTTACAGAGAGAACCTTCCCAAGGTGGGCTTCTGACTGTGGT 282

Qy 235 ATGCAGTCCAAAGTGTTCACGCGGTCTTCCACAGAGTGAAGGACTGTCTCACCATCAAC 294

Db 283 ATGCAGTCCAAAGTGTTCACGCGGTCTTCCACAGAGGAGGACTGTCTCACCATCAAC 342

Qy 295 GTGGTGGCGCGCGCGGCAACAGCGCGGCGCCAACTCCCGGTTCATGCTGTGATCTTT 354

Db 343 GTGGTGGCGCGCGCGGCAACAGCGCGGCGCCAACTCCCGGTTCATGCTGTGATCTTT 402

Qy 355 GCGGTGGGTGTGATCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 414

Db 403 GCGGTGGGTGTGATCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 462

Qy 415 AGTGTGCTCATGGGCAAG--CATCATCCAGTGGCGGTCAACTACCGTGTGCTCGTGG 473

Db 463 AGTGTGCTCATGGGCAAG--CATCATCCAGTGGCGGTCAACTACCGTGTGCTCGTGG 522

Qy 474 GGGTTCTTGGTGTGATGACATCAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 533

Db 523 GGTTCCTTGGCGGTCCGACATCAAGGCGGAGGCGAGCTCCAAATGCGCGGCTCAAGGAC 582

Qy 534 CAGCGTTTGGGCGATCGAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 593

Db 583 CAGCGTTTGGGCGATCGAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 642

Qy 594 AAGGTGA-CATCTTTGGGCGAG---GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 649

Db 643 AAGGTGA-CATCTTTGGGCGAGTCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 702

Qy 650 AACGACGGCGCAACACGTTAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 706

Db 703 AATGGCGCGCAACACGTTAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 762

Qy 707 GGAGCCATGGTGGCG---GACCCGCTGGACGCGACGTCAGCGCAACGAGATCTTACGACCTTC 763

Db 763 GGAGCCATGGTGGCGTCTGACCCGCTGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 822

Qy 764 TTTTGTCTCGAGTGTGGGTGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 820

Db 823 TTGGTGGCTTCTACGCGGCTGCAGCAGTGCACGCAACAGCTTGGCTGTGGCTGTGGCTGTCTT 882

474 GGGTTCTTGGCTGGTGATGATCAAGCGCGAGCGGAGCGGACCGCGCTTGAAGGAC 533
523 GGGTTCTTGGCTGGTGATGATCAAGCGCGAGCGGAGCGGAGCGGAGCGGCTTGAAGGAC 582
534 CAGCGTTTGGCGATGCGAGTGGTGCGGAGCAACATTCGCGGGTTTGGCGGCGAGCCCGAGC 593
583 CAACGCTTGGTTTGGTGGTGCGGAGCAACATTCGCGGGTTTGGCGGCGAGCCCGTCC 642
594 AAGGTGA-CATCTTTGGGGA---GGCGGGGAGCATGTCCTGGTTGTGCGCACCTCATCTGG 649
643 AAGGTGACCATCTTTGGTGGTGCGGCGGCGAGCATGTCGTAATGTGTGTCAGCTCCTCTGG 702
650 AACACGCGGAGCAACATGATCAAGGGCAAGCGGTTTTCGCGGCGGCGATCATGCGAG--- 706
703 AACACGCGGAGCAACATGATCAAGGGCAAGCGGTTTTCGCGGCGGCGATCATGCGAGTCT 762
707 GGAGCCATGTTGTC---CGGACCCCGTGGACGCGGACGTAACGCAACAGAGATCTAGACCTC 763
763 GGGGCCATGTTGCGGTGGACCCCGTGGATGGGCGCTACGGCAGCGAGATCTAGACCCAG 822
764 TTTGTCTCGAGTGTGCTGTGGAGCGCCAGCGAGCATGCTGGTGCTTGGCGAGTG-- 821
823 GTGGTTGCTTACGCGGGTGTGGAGTGGCGGACGAGTGGCGTGTGGCGAGCATC 882
822 -CGAGCGACCTTGTCTGATGCCACCAACACACTCTCTGGGTTCTTGGCGTACTCTCTCG 880
883 TCGAACGAGCAAACTCTTCCAGGCGCCAGCGGACGATCATCACCAGTACATGTTCAAG 942
881 TTGCGGTTGT-----ACTCCCGGCGGACGCGGAGCAACATCACCGATGATGATGACAG 934
943 TTGCGGTTGTCTTCTCCGCGGCGCGGACGAGCTTCTATCACCAGTACATGTTCAAG 1002
935 TTGGTGGCGGAGCGGAGTATGCAAGCTTCCGTTGATCATTCGGGCGGACGAGAACCGAG 994
1003 TTGGTGGCGGAGCGGAGTATGCAAGCTTCCGTTGATCATTCGGGCGGACGAGAACCGAG 1062
995 GGCACATCTTTGGCT-----CTTGAAGTGGACGAGGATGCTCAGGCGCGTCTTAC 1047
1063 GGCAGTGTGTGGCTTGTCCAGTTGATGATGATGATGATGATGATGATGATGATGATGAT 1122
1048 TTCAAGCA---GTTTCATCCAGCGGAGCGGAGATGCAACCTTGTATGCGGGGTAC 1104
1123 TTCAAGCAAGCTTTCATCCAGCGGAGCGGAGATGCAACCTTGTATGCGGGGTAC 1182
1105 CCGGAGGATCACCAGG---GTCCGTTGAGACGG-----TTCAAGCG---TCACCCCG 1154
1183 CCGAGGATCACCAGGAGTGTGCTTTCGACCGGATCTTCAACGCGGATCTTCAACGCGG 1242
1155 CAGTTCAAGAGAT---CGCGGTGTGCGGAGCTTGCATTCATCCAGCGGCGGCTAC 1211
1243 CAGTTCAAGAGATGTCAGCGGTGTGTTGATGATGATGATGATGATGATGATGATGATGAT 1302
1212 TTCTCTCAACCACTTCCAGGCGGCGACCAAGTACTCTTCTTCTCAAGCAGC-----TCGGG 1265
1303 TTCTCTCAACCACTTCCAGGCGGCGACCAAGTACTCTTCTTCTCAAGCAGCTTAGTGGG 1362
1266 TTGCCATCATGGGACCTTCCATGCAAGCAATGTTGCGGAGGACTACTTGTGGGA 1325
1363 TTGCCGTTGTTGGCACCCACCGCAAGCATTTGTGGGAGGACTTTTGTGGAGC 1422
1326 AGCGGAGGCTCATCTACAAACAGCGTTTATCGGTTGCGCAACCGACTTGGACCCCAAC 1385
1423 CACAGAGCGCGGTGTACAAACAGCGTTTATGCGTTTGTGCGCAACCGCTCGACCCGAC 1482
1386 ACAGCGGGGTGTGTGGTGAACCTGCGCCAAAGTACACGAGCAGC-----CAGGCGAACAAC 1439
1483 AAGGCGGTTTGTGTTGTAACCTGCGCCAAAGTACACGAGGCTCTCAGTCAGGCAACAAC 1542
1440 TTGATGATGATCAAGCGCTTGGCTGTACACGCGGAGGAGCAACTTTCGACCGCTGGC 1499
1543 TTGATGATGATCAAGCGCTTGGCTGTACACGCGGAGGAGCAACTTTCGACCGCTGGC 1602
1500 TAGACGCGTTGATGATCAACCGGTTCTTTGT 1531

1603 TAGACGCGTTGTTTACCAACCGCTCTCTTT 1634
RESULT 7
ABX95905
ID ABX95905 standard; DNA; 1641 BP.
XX AC ABX95905;
XX 15-JUL-2003 (first entry)
XX Candida rugosa lipase 2 DNA.
XX Lipase 2; gene; ds; mutant.
XX Candida rugosa.
XX
XX Key Location/Qualifiers
XX CDS 1..1641
XX /*tag= a
XX /product= "Lipase 2"
XX /partial
XX /note= "No start or stop codon shown"
XX
XX EP1288294-A2.
XX
XX 05-MAR-2003.
XX
XX 26-APR-2002; 2002EP-00009616.
XX
XX 31-AUG-2001; 2001US-00943857.
XX (SINI-) ACAD SINICA.
XX
XX Tang S, Lee G, Shaw J;
XX
XX WPI; 2003-395476/38.
XX P-PSDB; ABU09070.
XX
XX Isolated mutant nucleic acid encoding Candida rugosa lipase, useful for the preparation of Candida rugosa lipase for biocatalytic applications.
XX
XX Claim 25; Page 4-5; 33pp; English.
XX
XX The invention relates to an isolated mutant nucleic acid encoding a Candida rugosa lipase polypeptide. The DNA has a sequence having at least 80% identity to a wild-type DNA encoding Candida rugosa lipase. The microorganism is a bacterium or yeast, preparing a mutant DNA encoding a C. rugosa lipase and a chimeric C. rugosa lipase comprising a substrate interacting domain of a first C. rugosa lipase and a non-substrate interacting domain of a second C. rugosa lipase. The method is useful for preparing a mutant DNA encoding a Candida rugosa lipase. The DNA is useful in the large scale manufacture of Candida rugosa lipase which is useful for biocatalytic applications. This sequence represents DNA encoding Candida rugosa lipase 2
XX
XX Sequence 1641 BP; 305 A; 521 C; 472 G; 343 T; 0 U; 0 Other;
SQ
Query Match 57.0%; Score 873.2; DB 7; Length 1641;
Best Local Similarity 78.6%; Pred. No. 1.4e-180;
Matches 1252; Conservative 0; Mismatches 278; Indels 62; Gaps 15;
QY 2 CCACCGCAAGCTGCCAACGCGGACACCATCACCGGTCTCAACGCCATCATCAACGAG 61
Db 43 CCACCGCAAGCTGCCAACGCGGACACCATCACCGGTCTCAACGCCATTTGTCAACGAA 102
QY 62 GCGTCTCTCGGCAATCCCTTTGCGAGCGCGCGGTGGGCACTTCGCTTCAGGACCT 121
Db 103 AAGTTTCTCGGCATACGTTTGGCGAGCGCGCGGTGGGCACTTCGCTTCAGGACCT 162
QY 122 GTGCGGCTACTCTGGCTCGCTCAACGCGGAGAGTT----ACTTACGCGCGG---TGCGATG 174

163 GTCCGTTACTGGCGTCTGCTCAACGGCCAGAGTTTACCTCTTACGGCCGCTCTTGATG 222
 175 CAGCAGAACCCGAGGCGCAGTTTGAAGAGAACCTTGGCAAGACGGCACTCGACTTGGTG 234
 223 CAGATGAACCTTATGGGCTCGTTTGAAGACACATTTCCCAAGATGCGCTTGATTTGGTG 282
 235 ATGAGTCCAAAGTGTTCAGGCGGTGCTTCCCAAGAGTGAAGGACTGCTCACCATCAAC 294
 283 CTCAGTCCAAAGTCTTCCAAAGTGGTCTTCCCAAGACGAGGAGTGTCTCACCATCAAC 342
 295 GTGGTGGCGCGCGGCGGACCAAGCGGGGCGCCAACTCCCGCTCCCGCCAGATGGTCAACAA 414
 343 GTGATCCGGCGCGCGCGGACCAAGGCGGCGAGTGGTCTCCCGGCTGATGCTCTGATCTTT 402
 355 GGCGGTGGGTGTGAGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 414
 403 GGCGGTGGGTGTGAGCTTGGCGGCTCCAGGCTCTTTCCAGGAGACCAAGATGGTGGCCAG 462
 415 AGTGTGCTCANTGGGCAAGCC-ATCATCCAGTGGCGGTCAACTACCGTGTGGCTCGTGG 473
 463 AGCGTGTCTATGGGTAAACCGGTGATCCAGCTGAGCATGAATAGCGGTGGCGTCAATGG 522
 474 GGGTCTCTGGCTGATGATCAATCAAGCCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 533
 523 GGGTCTCTGGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 582
 534 CAGCGTTTGGCGATGAGTGGGTGGGAGACAAATTCGCGGCTTGGCGGCGGCGGCGGCGG 593
 583 CAGCGCTTGGCCATGAGTGGGTGGGCGGAGCAATTTGCTGGGTGGCGGCGGCGGCGGCG 642
 594 AGGTGA-CATCTTTGGCGAG---GGGGGAGCATGTCCGTGTGTGCGCACCTCATCTGG 649
 643 AAGTGAACCATATACGCGGAGTCTGGCGGCGAGCATGTTCGACGTTTGTGACCTTGTGTG 702
 650 AAGCGGGCGGCAACACGTACAGGCGGAGCGGCTTGTTCGCGCGGGCATCATGAGGGA 709
 703 AAGCGGGCGGCAACACGTACAGGCGGAGCGGCTTGTTCGCGCGGCCCATCATGAGTCT 762
 710 GCC-----ATGTTGCGCGGACCGGTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 763
 763 GGCTGATGTTGCGGCTGCTACCGGCTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 822
 764 TTTGTTCTCGAGTGTGGTGTGGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 820
 823 GTGGTGGCGTCTGCGGGGTGTGGCAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 882
 821 GCGAGCGGACCTTGTCTGATGCCACCAACACTCTCTGGGTCTTGGCGGCTACTCTCG 880
 883 TCTCAGGACACGTTGTACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 942
 881 TTGCGGTTGT-----ACTCCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 934
 943 TTGCGGTTGTCTTATCTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1002
 935 TTGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 994
 1003 TTGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1062
 995 GGCACCATCTTTGG-----CTCTTGAACGTGACCAAGATGCTCAGGCGCGGCGGCTTAC 1047
 1063 GGCATTTGTTTGGGCTCTCTTCTTTTGAACGTGACCAAGATGCTCAGGCGCGGCGGCTAC 1122
 1048 TTCAAGCAG----TTCAATCCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTAC 1104
 1123 TTCAAGCAGCTTTTATCCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTAC 1182
 1105 CCCCAGGACATACCCAGGG---TCCGTTTCAGACGG-----TTCAACGCTCACCCTGG 1154
 1183 ACCAGCGACATACCCAGGGTTCTCCGTTTCAGACCGGCGGATCTTCAATGCCATCACCCG 1242
 1155 CAGTTCAAGAGATC---GCGGTGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTAC 1211

Db 1243 CAGTTCAAACGAGTCTCTGCGTTGCTTGGCGACCTTGGCTTACGCTTGGCGTGGCTAC 1302
 Qy 1212 TTCTCAACACCTTCCAGGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1265
 Db 1303 TTCTCAACCTTACCTACGAGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1362
 Qy 1266 TTGCCAATCATGGGCGACCTTCCATGCGCAACGACATTTGTGGCAGGAGTACTTGTGGGA 1325
 Db 1363 TTGCCCGCTTGGCGACCTTCCAGCGGCAAGCATATCTTGGCAGGAGTACTTGTGGCG 1422
 Qy 1326 AGCGGAGCGCTCATCTACAAACAGCGGTTTATCGGTTCCGACCGACTTGGACCCCAAC 1385
 Db 1423 AGCGGAGTGTGATCTACAAACAGCGGTTTATCGGTTTGGCAACGACCTCGACCCGAC 1482
 Qy 1386 ACCGCGGGGTGTGGTGAATCTGGCCCAAGTACACGAGCAGC-----CAGGGCGAACAC 1439
 Db 1483 AAGCGGGGTGTGGGACCAACTGGCCCAAGTACACGAGCAGCTCTCAGTCTGGCAACAC 1542
 Qy 1440 TTGATGATGATCAACGCGCTTGGGCTTGTACACCGGCAAGGACAACTTCCGCGACCGCTGGC 1499
 Db 1543 TTGATGAGATCAACGCGCTTGGGCTTGTACACCGGCAAGGACAACTTCCGCGCGATGCG 1602
 Qy 1500 TAGCAGCGGTTGATGACCAACCGGTTCTTTGT 1531
 Db 1603 TAGCGCGCTCTTTTCCAACCGCGCTCTTT 1634

RESULT 8

AX33112
 ID AAX33112 standard; DNA; 1647 BP.

XX AAX33112;

XX 23-JUN-1999 (first entry)

XX Synthetic lipase 1 gene.

XX Candida rugosa; lipase 1; Lip1; industrial bioconversion; ss.

XX Synthetic.

XX Candida rugosa.

XX WO9914338-A1.

XX 25-MAR-1999.

XX 16-SEP-1997; 97WO-NL000524.

XX 16-SEP-1997; 97WO-NL000524.

XX (UNIL) UNILEVER NV.

XX Brocca S, Schmidt-Dannert C, Lotti M, Alberghina L, Schmid R;

XX WPI; 1999-229539/19.

XX Synthesis and functional overexpression of a Candida rugosa lipase gene coding for a major industrial lipase.

XX Disclosure; Page 23-28; 44pp; English.

CC The present sequence represents a synthetic lipase 1 gene derived from
 CC the native Candida rugosa lipase 1 gene. Lipases produced by Candida
 CC rugosa are extensively used in industrial bioconversions, and the pure
 CC lipase 1 can be used in a process requiring high specificity toward acyl
 CC chains shorter than 14C. Lipase 1, free of 2-5, can be obtained without
 CC using extensive and expensive working up procedures. Pure lipase 1
 CC exhibits higher activity toward caprinic than toward palmitic

XX Sequence 1647 BP; 377 A; 400 C; 364 G; 506 T; 0 U; 0 Other;

Query Match

42.5%; Score 651.8; DB 2; Length 1647;

Best Local Similarity

69.7%; Pred. No. 2.4e-132;

Matches 1104;	Conservative	0;	Mismatches 417;	Indels 62;	Gaps 14;
Qy	2	CCACCGCCAAAGCTCGCAACGGGACACCATCACCGGTCTCAACGCCATCATCAACGAG	61		
Db	49	CCAAACGCCAATTTGGCTTAACGGGTGACACCATCACCGGTCTCAACGCCATCATCAACGAA	108		
Qy	62	GGGTTCCTCGGCATCTCCCTTGGCGAGCCGCGGTGGCAACCTCCCGCTTCAAGGACCT	121		
Db	109	GCCTTCTTGGTATTCATTTGGCGAACCCACAGTTGGTAATTCAGATCAAGGACCA	168		
Qy	122	GTGCGGTACTCTGCTCGCTCAACGGCCAGAGTTACTTACGGCC-----CGTGCATG	174		
Db	169	GTTCATATCTCCGTTCTTGATGGTTCAGAGTTTCAAGTTTCACTTCTTACGGTCCATCTGTATG	228		
Qy	175	CAGCAGAACCCCGAGGACAGCTTTGAAGAGAACCTTGGCAAGACGCACTCGACTTGGTG	234		
Db	229	CAACAAACCCAGAGGTACCTACGAAAGAACTTGGCAAGAACTTGGCAAGGAGCTTTAGATCTGGTT	288		
Qy	235	ATGCAGTCCAGGTGTTCAGCGGTGTCTCCCGAGGTGAGGACTGCCTCACCATCAAC	294		
Db	289	ATGCAATCCAAAGTTTTCGAAGCTGTTTCTCCATCTCTGAGACTGTTGACCATTAAT	348		
Qy	295	GTGCTGCGCGCGCGGACCAAGCGCGCGCCCAACCTCCCGGTCTGCTCTGATCTTT	354		
Db	349	GTGTTAGACCAACCGGACAAAGGCTGTGCCAATTTGCCAGTTATGTTGTGGATCTTT	408		
Qy	355	GGCGGTGGGTTTGAATCGGAGGCCCAACATCTTCCCTCCCGCCAGATGTCACCAAG	414		
Db	409	GGTGGTGGTTTGAAGTTGGTGTACTAGTACTTCCCTCCAGCCCAATGATTACCAAG	468		
Qy	415	AGTGTCTCATGGGCAAGCC-ATCATCCAGTGGCGGTCAATACCGTGTTCCTCGCTGG	473		
Db	469	TCTATTGTATGGGTAAGCAATCATCAACGTTTCTGTCATCAAGACTCGAGCTGG	528		
Qy	474	GGGTTCTTGGTGTGTATGACATCAAGCGCGAGGCGAGCGGAAACCGCGGTGTAAGAAC	533		
Db	529	GGTTTCTTGGTGTGTGACGAATCAAGCGGAAGGTCTTGCCAAACGCGGTGTAAGAAC	588		
Qy	534	CAGCGTTGGGATGAGGTGGTGAGCAACATGTCGGGTTCGGCGGCGACCGGAGC	593		
Db	589	CAAGATTGGGTATGCAATGGGTGGCTGACAAATTTGCTGTCTTGGTGGTATCCAATCT	648		
Qy	594	AGGTGAC-ATCTTTGGCA---GGCGGCAGCATGTCCGTGTGTGCGCACCTCATCTGG	649		
Db	649	AAAGTTACTATCTTTGGTGAATCTGCTGTTCTATGTCCGTATGTGTGACATTTGGG	708		
Qy	650	AACGACGGCGCAACACGCTACAAAGGCAAGCGTTTCCCGCGCGGCATCANGCA--G	706		
Db	709	AACGACGGTGACACACTTCAAGGGTAAGCCATTGTTTCAGAGCTGGTATCATGCAATCT	768		
Qy	707	GGAGCCATGGTGC---CGGACCGGTGGAGCGGACGTCAGGCAACGAGATCTACGACCTC	763		
Db	769	GGTGTATGGTTCATCTGACGCGCTGACGGGTATCTACGGTAACGAAATTTTGACTTG	828		
Qy	764	TTTGTCTCGAGTCTGGCTGTGCGACGCGCAGCAAGCTCGCGTGTGGCGAGTG--	821		
Db	829	TTGGCTTCAACCGCTGGTGTGGTCTGCTCTGACAAAGTGGCTGTGTGGAGGTGTT	888		
Qy	822	-CGAGCGACACCTTGTCTGATGCCAACAAACACTCTCTGGGTCTTGGGCTACTCTCG	880		
Db	889	TCCTTCTGACACTTTGGAAGACGCCCAACAAACACCCCTGTTCTTGGCTTACTCTCC	948		
Qy	881	TTGCGGTTGTACT-----CCCGCGCGGACGCGCAAGAAATCACCGATGACATGTAAG	934		
Db	949	TTAAGATTGTCTTACTTGGCAAGACGAGCGGTGTTAATCATCACCGACGATGTACGCT	1008		
Qy	935	TTGGTGGCGACGCGAGTATGCAAGGTTCCCGTGTATCTTGGCGACCAAGACGAG	994		
Db	1009	TTGGTTAGGAAGTATGATGCAACATCCCTGTTATCATCGTGCACAAACGAGGAA	1068		
Qy	995	GGCACCATCTTTGGCTC-----TTGAACGTGACCAAGAAATGTCAGGCCCGGTCTTAC	1047		
Db	1069	GGTACCTTCTTTGGTACTTCTTCTTGAAGGTTTACCACTGATGTGCCAAGCCAGAGATAT	1128		

Qy	1048	TTCAAGCA---GTTTCATCCACGCGACGCGGAGATCGACACCTTGATGGCGCGCTAC	1104
Db	1129	TTCAAGCAATCTTTTGTCCACGCTAGCGACGCTGAATCGACACTTTTGATGACTGCTTAC	1188
Qy	1105	CCCCAGGACATCACCCAGGT-----CCGTTGCACACGTTCAACGCTCACCCCG	1154
Db	1189	CCAGGTGACATCACTCAAGGTTCTCCATTTGACACGTGAATTTAAAGCGCTTGACCCCA	1248
Qy	1155	CAGTTCAAGAGAAATC---GGGTGCTCGGCGACCTTGCTTCATTCACGCGCGCGCTAC	1211
Db	1249	CAATTCAGAGAAATCTCTGCTGTTTGGGTGACTTGGGTTTACTTTGGCTCGTAGATAC	1308
Qy	1212	TTCTCAACACCTTCCAGGCGGCAACCAAGTATCTGTTCTCTCAAGCAGCTC-----GGG	1265
Db	1309	TTCTTGAACCACTACACCGGTGGTACCAAGTACTCTTTCTTGTCTAAGCAATTTCTGGT	1368
Qy	1266	TTGCCAATCATGGCACCTTCCATGCCAAGCAATCTGTGGCAGGACTACTTGTGGGA	1325
Db	1369	TTGCCAGTTTGGGTACTTTCCACTCCCAACGATATCGTCTTCCAAAGACTACTTGTGGT	1428
Qy	1326	AGCGGACGCTCATCTACAAACACGCGTTTATCGCTTTCGCCACCGACTTGGACCCCAAC	1385
Db	1429	CTGCTTCTTGTATCTACAAACGCTTTCATTGCTTTTGCCACTGACTTGGACCCCAAC	1488
Qy	1386	ACCGCGGCTTGTGGTGAACCTGCCCAAGTACACC-----AGCAGCCAGGCGCAACAC	1439
Db	1489	ACCGCGGCTTGTGGTGAAGTGGCGCAAGATACACCTTCTTCTCAATCTGGTAACAC	1548
Qy	1440	TTGATGATGATCAAGCGCTTGGCTTGTACACCGCGAGGACAACTTCCGACCGCTGGC	1499
Db	1549	TTGATGATGATCAACGCTTGGGTGTTGTACACCGGTGAAGCAACTTCAGAACCGCGGT	1608
Qy	1500	TACGACGCTTGTATGACCAACCC	1522
Db	1609	TACGACGCTTGTCTCAACCC	1631

RESULT 9

ID	AAH43625	standard; cDNA; 1950 BP.
XX	AAH43625;	
AC	AAH43625;	
XX	21-JAN-2002 (first entry)	
DT		
XX	Synthetic lip1 gene.	
DE	Lipase; lip1; variant; ripening form; serine; 16-18C acyl chain; ss.	
XX	Candida rugosa.	
OS		
XX	Key	Location/Qualifiers
FT	CDS	175..1779
FT		/tag= a
FT		/product= "lip1"
FT		/transl_except= (pos:292..294,aa:Glu)
FT		/transl_except= (pos:301..303,aa:Gln)
FT		/transl_except= (pos:532..534,aa:Thr)
FT		/transl_except= (pos:553..555,aa:Thr)
FT		/transl_except= (pos:565..567,aa:Arg)
FT		/transl_except= (pos:802..807,aa:Thr Thr)
FT		/transl_except= (pos:1024..1026,aa:Glu)
FT		/transl_except= (pos:1114..1116,aa:Thr)
FT		/transl_except= (pos:1288..1290,aa:Thr)
FT		/transl_except= (pos:1678..1680,aa:Gly)
FT		/transl_except= (pos:1753..1755,aa:Phe)
XX		
PN	EP1130100-A1.	
XX		
PD	05-SEP-2001.	
XX		
PF	02-FEB-2001; 2001EP-00200375.	

XX PR 14-FEB-2000; 2000EP-00200513.

XX PA (UNIL) UNILEVER NV.

XX PA (UNIL) UNILEVER PLC.

XX PI Brocca S, Bornscheuer UT, Pleiss J, Schmid RD, Schmid U;

XX PI Schmitt J;

DR WPI; 2001-649825/75.

XX P-PSDB; AAB47622.

XX PT Modified lipolytic enzymes with altered substrate specificity, useful for

PT biocatalytic applications comprising high specificity towards carbon 16

PT and carbon 18 acyl chains.

XX PS Disclosure; Fig 1; 33pp; English.

XX CC This sequence encodes a parent lipase, lipl1, derived from C. rugosa. The

CC lipase of the invention is a variant of this parent lipase, with altered

CC properties. The variant is the ripening form of C.rugosa lipase selected

CC from pre, pro, prepri or mature lipase, in which 60% or less of the CNG

CC codons encoding serine in the native C.rugosa sequence, are replaced by a

CC universal codon for serine. The modified nucleic acid sequence is further

CC modified such that lipase variant exhibits an altered property. The

CC modified lipase is useful in a process requiring high specificity towards

CC 16-18C acyl chains

XX SQ Sequence 1950 BP; 472 A; 478 C; 420 G; 580 T; 0 U; 0 Other;

Query Match 42.5%; Score 651.8; DB 4; Length 1950;

Best Local Similarity 69.7%; Pred. No. 2.5e-132;

Matches 1104; Conservative 0; Mismatches 417; Indels 62; Gaps 14;

QY 2 CCCAACGCCAAGCTCGCAACGCGCACCATCACGGTCTCAACGCCATCATCAACGAG 61

DB 178 CCACCGCCACTTTGGCTAACGGTGACCATCACCGTTTGAACGCCATCATCAACGAA 237

QY 62 CGGTTCCTCGGCATTCCCTTTGCGAGCGCGGTGGGCAACCTCCGGTCTCAAGGACCCT 121

DB 238 GCCTTCTTGGGTATTCCAATTTGCCGAACACACAGTTGGTAATTCAGATTCAAGGACCCA 297

QY 122 GTCCGTACTCTGGCTCGCTCAACGGCGAGAAGTTACTTTAGCGCC-----CGTGAATG 174

DB 298 GTTCCATATCTCCGGTTCCTTGGATGGTCAAAGTTCACCTTCTTAGCGTCCATCTGTATG 357

QY 175 CAGCAGAACCCGAGGGACGTTTGAAGAGACCTTGGCAAGCGGACTCGACTTGGTG 234

DB 358 CAACAACCCAGAGGTACTTCAAGAAAACCTTGCCAAAGGCGAGCTTTAGATCTGGTT 417

QY 235 ATCAGTCCAAAGTGTTCAGAGCGGTGTCTCCCGAGAGTGAAGACTGGCTCACCATCAAC 294

DB 418 ATGCAATCCAAAGTTTTGCAAGCTGTTTCTCCATCTTCTGAAGACTGTTGACCATTAAT 477

QY 295 GTGGTCGGCGCGCGGACCAAGGGGGCGGCAACCTTCGGGTCACTCTGATCTTT 354

DB 478 GTTGTGTAGACCAACCGCGGACAAAGGCTGGTGCACATTTGCCAGTTATGTTGTGATCTTT 537

QY 355 GGCGGTGGTGTGAGATCGGAGCCCCACCATCTTCCCTCCCGCCAGATGGTCAACAG 414

DB 538 GGTGTGGTTTTGAAGTGTGGTACTAGTACTCTTCCCTCCAGGCCAAATGATTACCAAG 597

QY 415 AGTGTGCTCATGGGCAAGCC-ATCATCCAGTGGCGGTCAACTACCGTGTGCTCGTGG 473

DB 598 TCTATTGCTATGGGTAAGCCATCATCCAGTTTCTGTCACTACAGAGTCTCGAGCTGG 657

QY 474 GGGTCTTGGCTGGTATGACATCAAGGCGGAGGACGCGGAAACGCGGCTTTGAAGGAC 533

DB 658 GGTTCCTGGCTGGTGACGAAATCAAGGCGGAAGTCTCTGCCAACGCGGGTTTGAAGGAC 717

QY 534 CAGCGTTTCGGCATCGAGTGGGTGGCAGACAACATTTGCGGGTTTCGGCGGACCCGAGC 593

DB 718 CAAAGATTGGGTATGCAATGGGTGGCTGACAAACATTTGCTGTTTTGGTGGTGTATCCAACT 777

QY 594 AAGGTGAC-ATCTTTGSCGA---GGCGGCGACGATCTCGTGTGTGGCCACCTCATCTCG 649

DB 778 AAGGTACTATCTTTGGTGAATCTGCTGGTTCTATGTCGTCATGTGTGTCATCTTTGGTGG 837

QY 650 AACGACGGCGACAAACAGTACAGGCAAGCGGTGTTCCGCGCGGCGCATCATGCA---G 706

DB 838 AACGACGGTGAACAACATTTACAGGGTAAGCCATCTTTCAGAGCTGGTATCATGCAATCT 897

QY 707 GAGGCCATGCTGC---CGGACCCGGTGGAGCGGACGTCAGGCAAGAGATCTACAGACCTTC 763

DB 898 GGTGCTATGGTTCCATCTGACGCGGTTCGACGGTATCTACGGTAAGCAAAATTTTTCGCTTG 957

QY 764 TTGTGCTCGAGTCTGGCTGTGCGACGCGGACGCAAGCTCGGTGCTTGGCGAGTGG-- 821

DB 958 TTGGCTTCCACGCTGTTGTTGCTGCTCTGCAAGTTGGCTGTTTGTGAGAGGTGTT 1017

QY 822 -CGAGCGACACTTGTGCTGATGCCAACAAACACTCTCTGGGTGTTCTTGGCGTACTTCCTCG 880

DB 1018 TCTTCTGACACTTTTGAAGAGCGCCACCAACACCCCTGGTTCTTGGCTTACTCTCTCC 1077

QY 881 TTGGGTGTTGACT-----CCGCGCGGACGCGAAGAACATCAGCATGATGATGATAAG 934

DB 1078 TTAAGATTGTTTACTTTGCCAAGACGACGCGTGTAAACATCAGCGACGATGTACGCT 1133

QY 935 TTGGTGGCGGACGCGCAAGTATGCAAGCGTTCGCGTGTATCATTTGGCGACCCAGAACGCCAG 994

DB 1138 TTGGTTAGAGAGGTAAAGTATGCCAACATCCCTGTTTATCATCGGTGACCAAAACACGAA 1197

QY 995 GGCACCATCTTTGGCTC-----TTGAACGTGACACGAATGCTCAGGCCCGTCTTTAC 1047

DB 1198 GGTACCTCTTTGGTACTTCTTCTTTGAAGCTTACCCTGATGCCCAAGACGAGAAAT 1255

QY 1048 TTCAAGCA---GTTTATCAGCGCGGACGCGGAGATCGACCTTGTATGGCGCGGTAC 1104

DB 1258 TTCAAGCAATCTTTGTTCCACGCTAGCGACGCTGAATCGACACTTTGATGACTGCTTAC 1317

QY 1105 CCCAGGACATCACCCAGGTT-----CCGTTTCGACGCTTCAACGCTCACCCCG 1154

DB 1318 CAGGTGACATCATCAAGGTTCTCCATTTGACACTGGAATTTAAAGCCCTTGACCCCA 1377

QY 1155 CAGTTCAAGAGAAATC---GCGGTGCTCGGCGGACCTTGCAATTCATCCAGCGCGCGGTAC 1211

DB 1378 CAATTCAAGAGAAATCTCTGCTGTTTGGGTGACTTGGGTGTTTACTTTGGCTCGTAGATAC 1437

QY 1212 TTCCTCAACCATTTCCAGGCGGACCAAGTACTGTTCTCTCAAGCAGCTC-----GGG 1265

DB 1438 TTTCTTGAACCACTACACCGGTGTTACCAAGTACTCTTTCTTGTCTAAGCAATTTCTGGT 1497

QY 1266 TTGCCAATCATGGGACCTTCCATGCCAACGACATTTGTGTGGCAGGACTTCTGTTGGGA 1325

DB 1498 TTGCCAGTTTGGGTACTTTTCCATCTCAACAGGCTTTGCTTCTTCCAGACTTCTGTTGGGT 1557

QY 1326 AGCGGACGCTCATCTCAACAAACGCGTTTATTCGGGTTTCGCCACCGACTTGGACCCCAAC 1385

DB 1558 TCTGGTTCTTCTGATCTCAACAAAGCTTTTCACTGCTTTTGCCACTGACTTGGACCCCAAC 1617

QY 1386 ACCGCGGGTGTGGTGAAGTGGCCCAAGTACACC-----ACGAGCGAGGCAACCAAC 1439

DB 1618 ACCGCGGGTGTGGTGAAGTGGCCCAAGTACACCCTTCTTCTCAATCTGTTGTAACCAAC 1677

QY 1440 TTGATGATGATCAACGCTTGGGTGTTGATCGGCGGACGAGCAACTTTCCGACCGCTGGC 1499

DB 1678 TTGATGATGATCAACGCTTGGGTGTTGATACCCGGTAAGGACAACTTTCAGAACCGCGGT 1737

QY 1500 TACGACCGGTTGATGACCAACCC 1522

DB 1738 TACGACGCTTTGTTCTTCCAAACCC 1760

RESULT 10
AAO98578
ID AAO98578 standard; DNA: 2045 bp.

extracellular production of microbial aspartic protease. In particular, the enzyme aspergillopepsin is produced by transformed filamentous fungi and secreted into the culture medium due to the presence of a fungal secretion peptide. The present sequence is that of the A.foetidus SE4 glucoamylase promoter. (Updated on 16-OCT-2003 to standardise OS field)

SQ	Sequence	2045 BP	454 A	538 C	565 G	488 T	0 U	0 Other
Query Match	11.3%	Score 173	DB 2	Length 2045				
Best Local Similarity	56.6%	Pred. No. 5.2e-28						
Matches	426	Conservative	0	Mismatches	310	Indels	17	Gaps
5;								
QY	274	GAGAGCTGCCTCACCATCAAGTGGTGC	GGCGCGCGGACCAAGCGGGCCAACTC	333				
Db	70	GAGATTGCCCTGACATTTGACATTTG	CGCGTTCGCGCTCGCGGACCGCGGACTCG	129				
QY	334	CCGGTCATGCTCTGGATCTTTTGGCGGT	GGGTTTGAGATCGGCAGCCCAATCTTCCCT	393				
Db	130	CCTGTGCTGCTCGGATCTTTGGCGGAGG	CTTTGAACCTTGTTCAAGGCGATGATGAT	189				
QY	394	CCGCCCCAGATGGTACCAGAGTGTGCTC	TATGGGCAAGCC-AFATCCACGTGGCCGTC	452				
Db	190	GGTACACAGATGGTATCATCGTGCATAG	ACAAGAACATGCTATCGTGTGTTAGCAATG	249				
QY	453	AACTACCGTGTTCGCTCGTGGGGTTCCT	TCGGCTGGTGATGACATCAAGGCCGAGGCG	512				
Db	250	AATTATGCGGTGGAGGTTTCGGGTTCTT	GCGCCGAAAGAGATCTCTGAGGACGGGTCC	309				
QY	513	GGGAAACCGCGGCTTGAAGGACCAAGCG	TTTGGGCATGCAGTGGGTGGCAGACAACATG	572				
Db	310	GCGAACCTAAGGCTCTCGACCAACAGCC	TTCCTGCACTGGGTGCGCAACATCGAG	369				
QY	573	GGGTTCCGGGGCCACCGCAGCAGAGGTG	ACATCTTTGGCGAGGGCGGCGAGCATGTCG	632				
Db	370	GCCTTTGGTGGAGACCCGCAAGAGGTG	ACGATTTGGGGAGAATCAGCAGGA-GCCAT	428				
QY	633	TGTGCCACCTCATCTGGAAACGACGGCG	CAACACGTACAAGGGCAAGCCGTTGTTCCG	692				
Db	429	TTTGACTAGATGACTTGTACGACGGAA	ACATCACTTACAAGGATAGCCCTTGTTCGG	488				
QY	693	CGGGCATCATGACGGAGCCCATGGTGC	-----CGGACCGGGTGGACGGCACTGTA	746				
Db	489	GGGCCATCATGGACTCCGGTAGTGTGTT	CCCGCAGACCCCGTCGATGGGGTCAAGSG	548				
QY	747	ACGAGATCTACGACCTCTTTGTCTCGA	GTGCTGCTGTGCGCGCAGCGACGCAAGCTC	806				
Db	549	AGCAAGTATATGATCGGTAGTGGATCT	GAGGCTGTTCTCTTTAAAGCACCCCTAG	608				
QY	807	CGTGCTTGGCGAGTGGAGCGACAC---	CTTGTCTGATGCCACCAACAACACTCTTGG	863				
Db	609	CTTGTCTGGGTGAATAGACTACACCGA	CTTCCTCAATGCGGCAAACTCCGTGCCAGG	668				
QY	864	TCCTTGGCGTACTCTCGTTGGGTTGTAC	T-----CCCGGCCGACGGCAAGNATCA	917				
Db	669	TTTTAAGTACCAATCTGTGGCGGTATCA	TATGTGCTCTGACCGGACCGGCGTGTGT	728				
QY	918	CCGATGACATGTACAAGTTGGTGGCGA	CGGCAGATGCAAGCGTTCCTCGTGCATATTG	977				
Db	729	CGGCATCACCGACGTTTGGGCAAGCAG	GAGGAATATGCTCGGTCCTGTTATCGTGG	788				
QY	978	GGCACAGAACGACGAGGGCACCATCTT	TGGCT	1010				
Db	789	GGCACCAAGAGGATGAGGGGACCTTAT	TCGCT	821				

RESULT 12	
AAQ05605	
ID	AAQ05605 standard; cDNA; 1674 BP.
XX	
XX	
XX	AAQ05605;
XX	
DT	24-OCT-2003 (revised)
DT	25-MAR-2003 (revised)

DT	13-DEC-1990	(first entry)	
XX	Gene encoding protein with lipase activity.		
DE	Lipase.		
XX			
KW			
XX			
OS	Galactomyces geotrichum; ATCC34614.		
XX			
Key	Location/Qualifiers		
FH	40..1671		
FT	/*tag= a		
FT	/label= Claim 1		
XX			
FN	JP02174680-A.		
XX			
PD	06-JUL-1990.		
XX			
PF	27-DEC-1988; 88JP-00330598.		
XX			
PR	27-DEC-1988; 88JP-00330598.		
XX			
PA	(KURK) KURITA WATER IND LTD.		
PA	(OSAK) OSAKA CITY.		
XX			
DR	WPI; 1990-250686/33.		
XX			
DR	P-PSDB; AAR96370.		
XX			
PT	Protein gene having lipase activity - has defined sequence of 544 bases		
PT	and gives lipase producing vector on integration to expression		
PT	(secretion) vector.		
XX			
XX	Disclosure; Fig 4; 12pp; Japanese.		
XX			
CC	The cloned cDNA can be inserted into an expression vector and used to		
CC	transform hosts for the prodn. of a protein with lipase activity. See		
CC	also AAQ5606. (Updated on 25-MAR-2003 to correct PA field.) (Updated on		
CC	24-OCT-2003 to standardise OS field)		
XX			
SQ	Sequence 1674 BP; 351 A; 473 C; 381 G; 469 T; 0 U; 0 Other;		
	Query Match 11.2%; Score 171.2; DB 2; Length 1674;		
	Best Local Similarity 52.8%; Pred. No. 1.2e-27;		
	Matches 643; Conservative 0; Mismatches 508; Indels 67; Gaps 10;		
QY	270 GAGTGGAGGACTGCTCACCATCAACGTGGTGGCGCGCGCCGACCAAGCGCGCGCAA 329		
DB			
DB	342 GAATGAGGACTGCTTACCTCAATGTTTCCGCCCTGCTGGACCAAGCCTGATGCTAA 401		
QY	330 CTTCCCGGTCATGCTCTGGATCTTTGGCGGTGGGTTTGAGATCGGCAGCCCCACCATCTT 389		
DB			
DB	402 GCTCCCGTCATGGTTGGATTTACGGTGTGGGTTGTTTACGGTTCTTCTGCTGCCTA 461		
QY	390 CCTTCCCGCCAGTGGTACCAAGATGTGCTATGGGCAAG-CCATCATCCACGTGCG 448		
DB			
DB	462 CCCTGGTAAACAGTACGTTTAAAGAAAGTATCAACATGGGCGACCCCGCTTGTTGTTTC 521		
QY	449 CGTCAACTACCGTGTTCCTCGTGGGGTCTTTGGCTGGTGATGACATCAAGGCCGAGGG 508		
DB			
DB	522 CATCACTACCGTACCGGTCATTTGGATCTCTGGGTGGTGATGCCATCACCCTGAGGG 581		
QY	509 CAGCGGGAACGCCGCTTGAAGCACACGCTTTGGGCATGCACTGGGTGGCAGACAACAT 568		
DB			
DB	582 CAACACCAACGCTGGTCTGCACGACCAGCGAAGGCTCTCGAGTGGGTTAGCGACAACAT 641		
QY	569 TGGCGGGTTGCGCGGCGACCCGAGCAAGGTGA----CATCTTTGGCGAGCGCGGACGAT 624		
DB			
DB	642 TGCACAACTTTGGTGGTGATCCCGCATTAAGGTCATGATTTTCGGTGAGTCCGCTGGTCCCAT 701		
QY	625 GTCCGGTTGTGCCACTCATCTGGAAACGACGCGACACGTCATCAAGGCGAAGCGTT 684		
DB			
DB	702 GAGTGTGTCTCACCGCTTATTCGCTATGTTGGTGTGACAACTTCAACGGAAGAAGCT 761		
QY	685 GTTCCGCGGGGCATCATGCAG---GGAGCCATGGTGGCGGACCCGGTGGACGGCACGTA 741		

Db 762 TTTCACCTGCGCATCTTCAGTCTGGTGGCCCTCTTCTTACCAAGACTCTAGTCGGT 821
 Qy 742 CGGCAACGAGATCTACAGC-----CTCTTTGTCTCGAGTGTGCTGTGGCAG 789
 Db 822 TGGTCCCGATATTTCTTACACAGATTTGCTCAGTATGCCGATGTGACACTAGTGGCAG 881
 Qy 790 CGCCAGACGAGCTCGCGTCTTGGCAGTGGCAGCGACACCTTGTCTCGATGCCACAA 849
 Db 882 TGCCAAACGACACTCTGGAGTGTCTCCGACGAAGTCCAGCTCTGTCTTGACAGTCCCA 941
 Qy 850 CAACACTCTCGGGTTC-----TTGGGCTACTCC-----TGTGTGCGGTGTATCT 894
 Db 942 GAACTCTTACGATCTCAAGATCTGTTTGGTCTACTCCCTCAATTTCTTGGATTGGTCC 1001
 Qy 895 CGGGCCGACGCGAAGAACATCAGCATGACATGACAGTGTGGTGGCGGACGCAAGTA 954
 Db 1002 CAGACCCGACGCGCAACATTAATTCGCGATGCGCTTATGAGCTCTTCCGACGGGTAGATA 1061
 Qy 955 TGCAAGCGTTCGCGTATCATTTGGCGACAGAAACGAGGAGCACCATCTTTGGCTCT-- 1012
 Db 1062 CGCAAGTTCCTTACATTAGCGTAAACGAGATGAGGTACTGTCTTTGTCTCTGT 1121
 Qy 1013 -----TGAAGTGAACAGAAATGCTCAGCGCGGTCTTACTTCAAGCA---GTTCAATCA 1064
 Db 1122 TGCTCTCAACGCTACCAAGCTCCCATGTTAAGAGTGGTTGCAAGTACATTTTCTACGA 1181
 Qy 1065 CGCCAGACGCGGAGATGACACCTTGAATGGCGCGTACCCCGAGGACATCACCAGG 1124
 Db 1182 TGCTTCCGAGGCTTCCATTTGACCGTGTGTTGTGCTGTACCGGACGCTCTCTGTGG 1241
 Qy 1125 TCCG-----TTCCAGACGGTTCACGCTCACCCGCGATTCAGAGAATCGCGT 1174
 Db 1242 CTGCGCTTCCGCACTGGCATTTCTAATGCTGACCCCGCCAGTTCAAGCGTGTGGCG 1301
 Qy 1175 GCTCGGCACTTGCATTCATCCAGCGCGCGCTACTT-----CCTCAACA 1222
 Db 1302 CATCTTGTCCGATATGCTTTTCCAGTCTCCCGCGCGTGATGCTTAGCGCCACCAAGA 1361
 Qy 1223 CTTCCAGGCGGCGACCAAGTACTGTTCTCAGCAGTCTGGGTGGCAATCATGGCAC 1282
 Db 1362 CGTTAACCGCTGGACTTACCTTTGACCCCATCTGCACCAACCTGGTGCCATTTTGGGTAC 1421
 Qy 1283 CTTCCATCCCAACGATTTGTGTGCGAGTACTTGTGGAGCGGCGAGCGTCACTTA 1342
 Db 1422 TTTCCATGGCAAGAGTCTTCTTCAATTCATGTGACATTTGCGCGCTAACTCTCTA 1481
 Qy 1343 CAACAACGCTTTATCGGTTTCCGACCGACTTGAACCCCAACACCGGGGTTTGTGT 1402
 Db 1482 CTTCTGTTATTTTATTTCTTTGCGCAACCCACCATGACCCCTAATGTTGGTACTTAACCTGCT 1541
 Qy 1403 GAAGTGGCCCAAGTACAC 1420
 Db 1542 CCAGTGGGATCAATACAC 1559

RESULT 13

AAQ54020

ID AAQ54020 standard; DNA; 1828 BP.

XX AC

XX AC

XX 16-OCT-2003 (revised)

DT 25-MAR-2003 (revised)

DT 19-JUL-1994 (first entry)

XX XX

Lipase coding sequence of Geotrichum candidum.

XX Immobilisation; enzyme; cell wall; alpha agglutinin; AGA 1; FLO 1;

XX Major cell wall protein; glycosyl-phosphatidyl-inositol;

XX anchoring protein; alpha factor; alpha-agglutinin; invertase; inulinase;

XX alpha-amylase; Geotrichum candidum; lipase; enzymatic process;

XX fermentation; biodegradation; catalysis; ss.

XX OS Galactomyces geotrichum.
 XX Key Location/Qualifiers
 FH CDS 40..1731
 FT /*tag= a
 FT /product= "Lipase."
 FT sig_peptide 40..96
 FT /*tag= b
 FT mat_peptide 97..1728
 FT /*tag= C
 FT /product= "Lipase."
 XX WO9401567-A1.
 XX 20-JAN-1994.
 PD 07-JUL-1993; 93WO-EP001763.
 XX 08-JUL-1992; 92EP-00202080.
 PR 14-DEC-1992; 92EP-00203899.
 XX (UNIL) UNILEVER PLC.
 PA (UNIL) UNILEVER NV.
 PI Klis FM, Schreuder MP, Toschka H, Verrips CT;
 XX WPI, 1994-035071/04.
 DR P-PSDB; AAR47577.
 XX
 PT Immobilisation of enzymes to microbial cell wall - by prodn. of fusion
 protein of enzyme linked to anchoring protein.
 XX Claim 8; Page 49-52; 99pp; English.
 XX The lipase is used in a method to immobilise enzymes to a microbial cell
 wall. The coding sequence is used in the production of a recombinant
 polynucleotide which comprises a structural gene encoding a protein with
 catalytic activity (the lipase) and at least part of a gene encoding at
 least the C-terminus of a protein capable of anchoring in a eukaryotic or
 prokaryotic cell wall. The anchoring fragment or protein is selected from
 alpha agglutinin, AGA 1, FLO 1, major cell wall protein of lower
 eukaryotes or a proteinase of lactic acid bacteria. The recombinant
 polynucleotide preferably also comprises a sequence encoding a signal
 peptide to ensure secretion of the expressed product. The signal peptide
 is preferably derived from glycosyl-phosphatidyl-inositol, anchoring
 protein, alpha factor, alpha-agglutinin, invertase or inulinase, alpha-
 amylase of Bacillus or proteinases of lactic acid bacteria. The host
 microorganism can be used for performing enzymatic processes on an
 industrial scale. (Updated on 25-MAR-2003 to correct PN field.) (Updated
 on 16-OCT-2003 to standardise OS field)

XX Sequence 1828 BP; 389 A; 483 C; 419 G; 537 T; 0 U; 0 Other;

Query Match 10.6%; Score 162; DB 2; Length 1828;

Best Local Similarity 56.0%; Pred. No. 1.3e-25;

Matches 448; Conservative 0; Mismatches 315; Indels 37; Gaps 6;

Qy 248 TGTTCCAGCGGTGCTTCCCGAGTGAGGAGTCTCCATCAACGTTGCGCGCGC 307

Db 377 TGGCGCAGGTAGTGTCTCCATGAATGAGGAGTGTCTACCTTAACGTTTCCGCGCG 436

Qy 308 CGGGCACCAAGCGCGCGCCCAACCTCCCGGTCTGATCTGTGGATCTTGGCGGTGGTTTG 367

Db 437 CTGGCACCAAGCCTGATGCTAAGCTCCCGTTCATGGTTGATTTACGTTGCGGTG 496

Qy 368 AGATCGGAGCGCCCGACCATCTCCCTCCCGCCAGATGTCACAGAGTGTCATCG 427

Db 497 TGTTTGGTCTTCTGCTTCTTACCTGGTAACGGGTACGGGTACGAGAGTGTGGAATGG 556

Qy 428 GCAAGCC-ATCATCCACGTTGGCGGTCAACTACCGTGTTCCTCTCGTGGGGGTCTTGGCTG 486

Db 557 GCACCGCTTGTGTTGTTTTCATCACTACCGTACCGGCCCTCATGATGATCTTGGGTG 616

QY 487 GTGATGATCAAGCCGAGGCGAGCGGGAACGCGCGCTTGAAGGACGAGCGTTGGGCA 546
 |||||
 Db 617 GTGATGCCATCACCCTGAGGCGCAACCAACGCTGGTCTGCAAGGCGGCGAGGGTC 676
 |||||
 QY 547 TGCAGTGGTGGCAGACAAATTCGCGGGTTCGCGGGGACCGGAGCAAGTGA-----CA 602
 |||||
 Db 677 TCGAGTGGGTAGGCGCAACATTCGCACTTTGGTGGTGAATCCCGCAAGGTCATGATT 736
 |||||
 QY 603 TCTTTGGCGGGCGGAGCATGTCGTTGGTGGCCACCTCATCTGGAACGACGCGGACA 662
 |||||
 Db 737 TCGTGTAGTCCGCTGGTGCCAGAGTGTGCTTACCAGCTTGTGGCTACGGTGGTGACA 796
 |||||
 QY 663 ACAGCTAAGGGCAAGCGTGTTCGCGCGGGCATCATCGAGGGAG----- 710
 |||||
 Db 797 ACACCTCAACGGAAGCAGCTTTTCCACTCTGCCATCTTCAGTCTGGCGGTCTCTTC 856
 |||||
 QY 711 ---CCATGGTCCGGACCGGTGACGCGCAGTACGGGAAGCATCTACGACCTTTTG 767
 |||||
 Db 857 CTTACTTTGACTCTACTTCTGTTGGTCCCGAGAGTGGCTACAGCAGATTTGCTCAGTATG 916
 |||||
 QY 768 TCTCAGTGTGGCTGTGGCAGCGCCAGCGACAGCTCGGCTGCTTGGCAG---TGGCA 824
 |||||
 Db 917 CCGATGTGACACAGTGCAGTGATGATGACACTCTGGCTTGTCTCCGAGCAAGTCCA 976
 |||||
 QY 825 GCGACACTTCTCGATGCCAACAAC-----AACACTCCTGGGTTCTTGGCG 871
 |||||
 Db 977 GCGATGTCTTGACAGTGGCAGAACTCGTATGATCTTAAGGACCTGTTGGTCTCTCC 1036
 |||||
 QY 872 TACTCTCTGTGGTGTACTCCGCGCGGCGAGCGGAAGCAATCACCGATGACATGATAC 931
 |||||
 Db 1037 CTCAATCTCTTGGATTGGTGC-CCAGACCGGACGCAACATTTATCCCGATGCCGCTTAT 1095
 |||||
 QY 932 AAGTTGTGTGCGGAGCGCAAGTATGCAAGCTTCCCGTGTATCTTGGCGGACGACGAC 991
 |||||
 Db 1096 GAGCTCTACGCGGAGTATGACGCAAGGTTCCCTACATTTACTGGCAACCGAGGAT 1155
 |||||
 QY 992 GAGGCGACCATCTTTGGCTC 1011
 |||||
 Db 1156 GAGGCTACTATTCTTGCCCC 1175
 |||||

RESULT 14
 AAT10421
 ID AAT10421 standard; DNA; 1635 BP.
 XX AC AAT10421;
 XX AC AAT10421;
 DT 16-OCT-2003 (revised)
 DT 20-SEP-1996 (first entry)
 XX Geotrichum candidum lipase gene.
 DE
 XX
 KW Fatty acid; transgenic plant; exogenous; lipase; Rhizopus niveus;
 KW Pseudomonas aeruginosa; Pseudomonas fluorescens; Geotrichum candidum;
 KW Candida cylindracea; constitutive; tissue specific; promoter; lipid;
 KW milling; biofuel; lubricant; detergent; ss.
 XX
 OS Galactomyces geotrichum.
 XX
 XX
 PH Key Location/Qualifiers
 FT 1..1635
 FT /*tag= a
 FT /product= "lipase protein"
 XX
 PN FR2722798-A1.
 XX
 PD 26-JAN-1996.
 XX
 PF 25-JUL-1994; 94FR-00009272.
 XX
 PR 25-JUL-1994; 94FR-00009272.
 XX

(NAPO-) INST NAT POLYTECHNIQUE TOULOUSE.
 Alibert G, Mouloungui Z, Boudet A;
 WPI; 1996-107680/12.
 Prodn. of fatty acids or derivs. from transgenic oilseed plants -
 engineered to express a lipase that contacts lipid(s) only when seeds are
 milled.
 Claim 9; Page 25-26; 32pp; French.
 A novel method of producing fatty acids or their derivs. in a plant
 comprises generating a transgenic plant contg. an exogenous lipase gene,
 esp. selected from the Rhizopus niveus, Pseudomonas aeruginosa,
 P.fluorescens, Pseudomonas sp., Geotrichum candidum or Candida
 cylindracea lipase genes (AAT10417-22 resp.). The lipase genes can be
 placed under control of a constitutive or a tissue specific promoter. The
 production of the fatty acids only occurs when the lipase and lipids
 contact each other after milling of the plants. The fatty acids generated
 can be used to prod. e.g. biofuels, lubricants, detergents, etc. (Updated
 on 16-OCT-2003 to standardise OS field)
 Sequence 1635 BP; 323 A; 448 C; 395 G; 469 T; 0 U; 0 Other;
 Query Match 10.5%; Score 160.2; DB 2; Length 1635;
 Best Local Similarity 55.8%; Pred. No. 3e-25;
 Matches 446; Conservative 0; Mismatches 318; Indels 35; Gaps 6;
 QY 248 TGTTCAGGCGGTCTTCCCGAGGTGAGGACTGCCTCACCATCAACGTTGGCGCGC 307
 |||||
 Db 281 TGGCCCGAGGTAGTGTCTCCATGAATGAGGACTGTCTTACCTTAAGTTTCCGCCCTG 340
 |||||
 QY 308 CGGGACCAAGGCGGCGGCAACCTCCCGGTATGCTCTCTGGATCTTTGGCGGTGGTTG 367
 |||||
 Db 341 GTGGCACCAAGCGCTGATGCTAAGCTCCCGCTCATGGTTTGGATTTTACGGTGGCTTTG 400
 |||||
 QY 368 AGATCGGCGAGCCCAACATCTTCCCTCCCGCCAGATGGTCCACAGAGTGTCTCATGG 427
 |||||
 Db 401 TGTGTGTTCTTCTGCTTCTTACCTTGTAACTGAGGACTGACGAGGAGTGTGAAATGG 460
 |||||
 QY 428 GCAAGCC-AATCATCCAGTGGCGGTCAACTACCGTGTTCCTGCTGGGGGTCTTGGCTG 486
 |||||
 Db 461 GCCAGCTGTGTGTTGTTTCCATCAACTACCTACCGGCGCCCTATGATTCCTGGTG 520
 |||||
 QY 487 GTGATGATCAAGCGGCGGAGCGGGAACCCCGGCTTTGAAGGACCGAGCTTTGGCA 546
 |||||
 Db 521 GTGATGCCATCACCGCTGAGGGTAAACCAACGCTGTCTGACAGCACCGCAAGGGTC 580
 |||||
 QY 547 TGCAGTGGGTGGCAGACAACTTGGCGGTTCGGCGGCGACCCGAGCAAGGTCA---CA 602
 |||||
 Db 581 TCGAGTGGGTAGCGAACAATTTGCCAATTTGGTGTGATCCCGACAAGGTCAATTT 640
 |||||
 QY 603 TCTTTGGCGAGGCGGCGAGCATGCTCGTGTGTGCCACCTCATCTGGAACGAGCGGACA 662
 |||||
 Db 641 TCGGTGAGTCCGCTGGTGGCATGAGTGTGCTCACCAAGCTTTGCTCTACGGTGGTGACA 700
 |||||
 QY 663 ACAGGTACAGGCGCAAGCGTGTTCGCGCGGGGATCATGAGGAG----- 710
 |||||
 Db 701 ACACCTACAAACGGAAGAGCTTTTCCACTCTGCCATTTCTCAGTCTGGCGGTCTCTTC 760
 |||||
 QY 711 ---CCATGTGCGGACCGGCGGAGCGGACGCTACCGGCAAGAGATCTACGACCTTTG 767
 |||||
 Db 761 CTTACTTTGACTCTACTTCTGTTGGTCCCGAGAGTGCCTACAGCAATTTGCTCAGTATG 820
 |||||
 QY 768 TCTCGAGTCTGGCTGTGGCAGCGCGCAGCAAGCTCGCGTCTTGGCGAG---TGGCA 824
 |||||
 Db 821 CCGGATGTGATGCCAGCGCAGTGACAATGAATCTGTGCTTGTCTCCGAGCAAGTCCA 880
 |||||
 QY 825 GCGACACTTGTCTGATGCGCAACAAC-----AACACTCCTGGGTTCTTGGGTTACTCC 877
 |||||
 Db 881 GCGATGTCTTGACAGTGGCCAGAACTCGTAGGATCTCAAGGACCTGTTTGGCTCTCTCC 940
 |||||

QY 878 TCGTTGCGTTGTACT-----CCGGCCCGGACGCAAGACATCAACCGATGACATGTACA 932
 Db 941 CTCATTCCTTGGATTTGGTCCCGAGACCCGACGCAACATTTTCCCGATCCCGCTTATG 1000
 QY 933 AGTTGGTGGCGGCGGCAAGTATGCAAGCGTTCCCGTGATCATTTGGCGACCAAGACGACG 992
 Db 1001 AGCTTACCGGACGCGGTAGATACGCAAGGTTCCCTACATTTACTTGGTAAACGAGGAGTATG 1060
 QY 993 AGGCGACCATCTTTGGCTC 1011
 Db 1061 AGGCTACTATTCTTGCCCC 1079

RESULT 15

AAQ10313
 ID AAQ10313 standard; DNA; 1692 BP.

XX AC AAQ10313;

XX DT 24-OCT-2003 (revised)

XX DT 25-MAR-2003 (revised)

XX DT 05-APR-1991 (first entry)

XX Sequence encoding protein with lipase activity.

XX KW ATCC 34614; ds.

XX OS Galactomyces geotrichum.

XX FH Key Location/Qualifiers

XX FT CDS 1..1692

XX FT /*tag= a

XX JF JP02299588-A.

XX PD 11-DEC-1990.

XX PF 27-MAR-1989; 89JP-00074721.

XX PR 27-MAR-1989; 89JP-00074721.

XX PA (KURK) KURITA WATER IND LTD.

XX PA (OSAKA) OSAKA CITY.

XX WPI; 1991-027567/04.

XX P-PSDB; AARI0330.

XX Gene for coding protein with lipase activity - is prepd. from messenger ribonucleic acid of geo-trichum candidum ATCC 34614.

XX Claim 1; Fig 4; 12pp; Japanese.

XX The gene product may be isolated from a transformed expression sytem, and CC may be enhanced with stability in heat, alkalai, acid and organic solvent CC by position-specific modulation. (Updated on 25-MAR-2003 to correct PA CC field.) (Updated on 24-OCT-2003 to standardise OS field)

XX Sequence 1692 BP; 336 A; 470 C; 399 G; 487 T; 0 U; 0 Other;

Query Match 10.3%; Score 157.8; DB 2; Length 1692;

Best Local Similarity 54.1%; Pred. No. 1e-24;

Matches 432; Conservative 0; Mismatches 332; Indels 35; Gaps 4;

QY 248 TGTTCAGCGGTGCTTCCCGAGTGGAGTACTGCTCACCATCAACGTGGTGGCGCGC 307

Db 338 TGGCCCGAGGTAGTGTTCATGATGAGGACTGTCTTACCTTAACGTTTTCCGCCCG 397

QY 308 CGGGCCACCAAGGCGGCGCAACTCCCGTCAATGCTCTGGATCTTTGGCGTGGTTTG 367

Db 398 CTGGCACCAGCTGATGTCTAAGTCCCGTCAATGTTTGGATTTACGGTGGTGGCTTTG 457

QY 368 AGATCGCGACGCCCAACATCTTCCCTCCCGCCAGATGGTCACAGAGTGTCTATGG 427

|||||

|||||

|||||

|||||

|||||

|||||

Db 458 TGTTTGTGTTCTTGTCTTTTAACTCGGTAAAGCGCTACGTCAAGAGAGTCTGGAATGG 517
 QY 428 GCAAGCC-ATCATCCAGTGGCCCTCAACTACCGTGTGGCTTGGCTTGGGGTCTTGGGTG 486
 Db 518 GCCAGCTGTGTGTTGTTTCCATCAACTACCGTACCGGCCCTATGATTCCTGGGTG 577
 QY 487 GTGATGACATCAAGCCGAGGCGAGCGGGAACCGCGCTTGAAGGACCAAGCTTTGGCA 546
 Db 578 GTGATGCCATCACCGCTGAGGGTAACACCAACCGTGTCTGACGACCAAGCGGATC 637
 QY 547 TGCAGTGGTGGCAGACAAACATTCGCGGTTTCGCGGCGACCCGAGCAAGTGA-----CA 602
 Db 638 TCGAGTGGGTAGCGCAACATTTGCCAACTTTGGTGTGATCCCGACAGATCATGATTT 697
 QY 603 TCTTTGGCGAGGCGGCGAGCATGTCGTTGTGTCACCTCATCTGGAAACAGCGGACA 662
 Db 698 TCGGTGAGTCCGCTGGTGGCCATGAGTGTGCTCACCAGCTTGTGCTTATGTTGGTGA 757
 QY 663 ACAGTACAGGGCAAGCCGTTGTTCCGCGGCGGATCATGACAGGAG-----710
 Db 758 ACACATACACAGAAAGCAGCTTTTCCACTCTGCCAATCTTCAGTCTGGCGTCTCTTC 817
 QY 711 ---CCATGTTGCGCGACCCGCTGGACGCGACGTACGCGCAACGAGATCTACGACCTCTTTG 767
 Db 818 CTTACTTTGACTCTACTTCTGTTGTTGTCGAGAGTGCCTACAGCAGATTTGCTCAGTATG 877
 QY 768 TCTCGAGTGTGCTGTGGCAGCGCCAGCAAGCTCGCGTGTGGCGAGTGGCGAGG 827
 Db 878 CCGGATGTGATGCCAGCGCGGTGACAAATGAACTCTGGCTTGTCTCCGACCAAGTCCA 937
 QY 828 ACACCTTGTGATGCCACCAACACACTCTCTGGGTTCTTGGCGTACTCTCTGTTGC--- 884
 Db 938 GCGATGCTTGACACAGTGCACAGAACTCGTACGATCTCAAGGACCTGTTTGGTCTGCTCC 997
 QY 885 -----GTTGTACTCCCGCGCCCGACGCGCAAGAACATCAACCGATGACATGTACA 932
 Db 998 CTCAAATTCCTTGGATTTGGTCCAGACCCGACGCGCAACATTTATCCCGATGCGCTTATG 1057
 QY 933 AGTTGGTGGCGACCGCAAGTATGCAAGCGTTCCCGTGTATCATTTGGCGACCAAGACGAG 992
 Db 1058 AACTCTACCGCAGCGGTAGATACGCGCAAGGTTCCCTACATTAAGTAAACAGGAGGATG 1117
 QY 993 AGGCGACCATCTTTGGCTC 1011
 Db 1118 AGGGTACTATTCTTGCCCC 1136

Search completed: August 5, 2004, 16:02:13

Job time : 476 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 2004, 15:47:40 ; Search time 95 Seconds
(without alignments)
8949.313 Million cell updates/sec

Title: US-09-943-857-3
Perfect score: 1532
Sequence: 1 gccaccgcaagctgcga.....tgaccaaccgcttttgg 1532

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 582709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1022	66.7	1650	2	US-08-776-210-6
2	173	11.3	2045	1	US-08-379-926A-7
3	164.2	10.7	1828	3	US-08-362-525-11
4	161.8	10.6	1635	2	US-08-776-210-5
5	88.2	5.8	1738	2	US-08-932-376A-1
6	87	5.7	8533	1	US-07-846-181-6
7	87	5.7	8533	1	US-07-845-989-6
8	61.2	4.0	2184	1	US-08-445-050-8
9	61.2	4.0	2184	1	US-08-204-691-1
10	61.2	4.0	2428	1	US-08-445-050-8
11	61.2	4.0	2428	1	US-08-204-691-1
12	61.2	4.0	2428	4	US-09-355-295B-2
13	61.2	4.0	2487	3	US-08-370-223-12
14	60	3.9	1845	1	US-07-732-962A-1
15	60	3.9	1845	5	PCI-US92-06106-1
16	60	3.9	2256	2	US-08-318-826A-5
17	60	3.9	2256	2	US-08-370-156-1
18	60	3.9	2256	3	US-08-814-095-1
19	60	3.9	3016	2	US-08-318-826A-7
20	60	3.9	3016	2	US-08-370-156-5
21	60	3.9	3016	3	US-08-814-095-5
22	60	3.9	3096	2	US-08-318-826A-6
23	60	3.9	3096	2	US-08-370-156-3
24	60	3.9	3096	3	US-08-814-095-3
25	60	3.9	35060	3	US-08-814-095-7
26	59.6	3.9	2344	4	US-09-347-878-31
27	59.6	3.9	3018	1	US-08-347-718B-3

ALIGNMENTS

RESULT 1

US-08-776-210-6
; Sequence 6, Application US/08776210
; Patent No. 5942659

; GENERAL INFORMATION:

; APPLICANT: ALBERT, Gilbert

; APPLICANT: MOULOUNGUI, Zephirin

; APPLICANT: BOUDET, Alain

; TITLE OF INVENTION: PROCESS FOR PRODUCING FATTY ACIDS OR

; TITLE OF INVENTION: DERIVATIVES THEREOF FROM OLEAGINOUS PLANTS

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: YOUNG & THOMPSON

; STREET: 745 South 23rd Street

; CITY: Arlington

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/776,210

; FILING DATE: 24-JAN-1997

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: FR 94 09272

; FILING DATE: 25-JUL-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: WO PCT/FR95/00957

; FILING DATE: 18-JUL-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: PATCH, Andrew J

; REGISTRATION NUMBER: 32,925

; REFERENCE/DOCKET NUMBER: IN 387 - BE 6996

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-521-2297

; TELEFAX: 703-585-0573

; TELEX: 248425 EMBON

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1650 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

US-08-776-210-6

Sequence 3, Appli
Patent No. 5200183
Sequence 3, Appli
Sequence 33, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 3, Appli
Sequence 1, Appli
Sequence 2, Appli
Sequence 1, Appli
Sequence 2, Appli
Sequence 11, Appli
Sequence 11, Appli
Sequence 11, Appli
Sequence 11, Appli
Sequence 11, Appli
Sequence 11, Appli

28 59.6 3.9 3018 1 US-08-482-262-3
29 59.6 3.9 3018 6 5200183-1
30 59 3.9 2220 2 US-08-932-376A-3
31 56.6 3.7 1905 4 US-09-347-878-33
32 56.6 3.7 1907 1 US-08-462-884A-2
33 56.6 3.7 1908 1 US-08-461-881B-2
34 56.6 3.7 1908 2 US-09-123-960-2
35 51.2 3.3 1746 4 US-10-023-515-3
36 51.2 3.3 2158 4 US-10-023-515-1
37 49 3.2 4403765 3 US-09-103-840A-2
38 49 3.2 4411529 3 US-09-103-840A-1
39 47.8 3.1 1590 4 US-09-351-150A-2
40 47.8 3.1 1800 1 US-08-484-815-11
41 47.8 3.1 1800 3 US-08-888-949-11
42 47.8 3.1 1800 3 US-08-888-950-11
43 47.8 3.1 1800 3 US-09-262-758-11
44 47.8 3.1 1800 4 US-09-885-876-11
45 47.8 3.1 1800 4 US-09-885-901-11

Query Match 66.7%; Score 1022; DB 2; Length 1650;
Best Local Similarity 84.7%; Pred. No. 2e-241;
Matches 1342; Conservative 0; Mismatches 180; Indels 62; Gaps 15;

QY 2 CCCACGCCAAGCTCGCAACGGCGACACCATCATCCGGTCTCAACGGCATCATCAACGAG 61
DB 49 CCCACGCCAAGCTCGCAACGGCGACACCATCATCCGGTCTCAACGGCATCATCAACGAG 108
QY 62 GCGTTCCTCGCATTCCTCTTTCGAGCGCGCGGTGGGCAACCTCCGCTTCAAGGACCT 121
DB 109 GCGTTCCTCGCATTCCTCTTTCGAGCGCGCGGTGGGCAACCTCCGCTTCAAGGACCT 168
QY 122 GTGCGGTACTCTGCTCGCTCAACGGCGCAGAGTT-----ACTTACGGCCCGTGCATG 174
DB 169 GTGCGGTACTCTGCTCGCTCGATGGCCAGAGTTTACGCTGTACGGCCCGCTGTGATG 228
QY 175 CAGCAGAACCCCGAGGGGACACCTTTGAAGAGAACCTTGGCAAGCGGCACTCGATTTGGT 234
DB 229 CAGCAGAACCCCGAGGGGACACCTTACGAGGAGAACCTTCCCAAGGCGAGCGCTCGATTTGGT 288
QY 235 ATGAGTCCAAAGTGTTCAGCGGTGCTTCCAGAGTGGAGTGGCTCACCATCAAC 294
DB 289 ATGAGTCCAAAGTGTTCAGCGGTGCTTCCAGAGTGGAGTGGCTCACCATCAAC 348
QY 295 GTGCTGCGCGCGCGGACCAAGCGCGGCGCAACCTCCCGGTCTGATCTCTGATCTTT 354
DB 349 GTGCTGCGCGCGCGGACCAAGCGCGGTGGCAACCTTCCCGGTGATCTCTGATCTTT 408
QY 355 GCGGTGCGGTGTGAGATCGGAGCGCCACCATCTTCCCTCCCGCGGAGATGTCACCAAG 414
DB 409 GCGCGCGGTGTGAGTGGTGGCACCAGCACCTTCCCTCCCGCGGAGATGTCACCAAG 468
QY 415 AGTGTGCTATGGGCAAG-CCATCATCGATGGCGGTCACTACGTTGTCCTCGTGG 473
DB 469 AGCATTTGCGCGCGGCGGACCAAGCGGTGGCAGTGGAGTGGTGGTGGTGGTGG 528
QY 474 GGGTTCCTGGTGTGATGACATCAAGCGCGAGGCGAGCGGAAACCGCGGTGGAAGGAC 533
DB 529 GGGTTCCTGGTGTGATGACATCAAGCGCGAGGCGAGTGGAGTGGTGGTGGTGGTGG 588
QY 534 CAGCGTTCGGGATGCGAGTGGTGGGAGCAACATTTGGCGGTTCGGCGGCGACCGGAGC 593
DB 589 CAGCGTTCGGGATGCGAGTGGTGGGAGCAACATTTGGCGGTTCGGCGGCGACCGGAGC 648
QY 594 AAGGTGA-CATCTTTGGGCA-----GGCGGCGAGTGTGCTGTTGTCACCTCATCTGG 649
DB 649 AAGGTGACCATCTTTGGCGAGTGGCGGCGAGTGTGCTGTTGTCACCTCATCTGG 708
QY 650 AACGAGCGGCAACACAGTACAGGCGAAGCGGTGTTCCGCGCGGCGATCATGCA--G 706
DB 709 AACGAGCGGCAACACAGTACAGGCGAAGCGGTGTTCCGCGCGGCGATCATGCAAGT 768
QY 707 GGAGCCATGTTGCC---GGACCGGTGGAGCGGACGTAGCGGAAAGAGATCTACGACCTC 763
DB 769 GGGGCCATGTTGCCCGGTGAGCGCGGTGGAGCGGATCTAGCGGAAAGAGATCTTACCTC 828
QY 764 TTGTCTCAGTGTGCTGTGTCAGCGCGGCGAGCGGCAAGCTCGGTTCGGCGAGTGC- 822
DB 829 TTGGCGTCAACCGCGGCTGCGGCGAGCGGCGAGCGGCAAGCTTCGGGTTCGGCGGTG 888
QY 823 --GAGCGACACCTTGTCTGATGCCAACACACATCTCTGGGTTCCTGGGTACTCTCG 880
DB 889 CTGAGCGACAGTGTGAGGAGCGGCAACACACACCTCTGGGTTCCTGGGTACTCTCG 948
QY 881 TTGCGGTG-----TACTCCCGCGCGGCGGCAAGAACATCACCATGATGATGATCAAG 934
DB 949 TTGCGGTGCTGTACCTCCCGCGGCGAGCGGTGAACATCACCGACGACATGTACGCC 1008
QY 935 TTGTCGCGGACCGGAGTATGCAAGGTTCCGCTGATCATTTGGCGGCGAGGACGAG 994
DB 1009 TTGTCGCGGAGGCAAGTATGCAACATCCCTGTGATCATTCGGCGGCGAGGACGAG 1068
QY 995 GGCAACCATCTTG-----GCTTTGAAGTGAACCGACCAATGCTCAGGCGCGGTGCTTAC 1047

DB 1069 GGCACCTTCTTTGGCACCTGCTGTGAACGTGACCGGATGCCAGGCCCGGAGTAC 1128
QY 1048 TTCAAGCA---GTTTCATCCACGCGCAGGAGATCGACACTTGTGATGGCGCGTAC 1104
DB 1129 TTCAAGCAGCTGTTTTCACGCGCAGGAGATCGACACTTGTGATGAGCGCGTAC 1188
QY 1105 CCCAGAGACATCACCCAGGG---TCCGTTTCGACACGGTT-----CAACGCTCACCCCG 1154
DB 1189 CCGGCGACATCACCCAGGGCTGCGCTTCGACACGGGTATTTCTCAACGCGCTCACCCCG 1248
QY 1155 CAGTTCAAGAGATC---GCGGTCTCGGAGACCTTGCATTCATCCAGCGCGCGCTAC 1211
DB 1249 CAGTTCAAGAGATCCTGCGGTGCTCGGAGACCTTGGCTTTACGCTTCTCGCTGCTAC 1308
QY 1212 TTCTCTCAACCACTTCCAGGGCGGACCAAGTACTGTTCTCTC-----AAGCAGCTCGGG 1265
DB 1309 TTCTCTCAACCACTACACGCGGCGGACCAAGTACTCATTTCTCTCTGAAGCAGCTCTCTGGC 1368
QY 1266 TTGCAATCATGGGACCTTTCATGCGCAACGACATTTGTGTGGAGGACTACTTGTGGGA 1325
DB 1369 TTGCGGTGCTCGGAACTTCCACTCCAAACGACATTTGTTCGAGGACTACTTGTGGG 1428
QY 1326 AGCGGAGCGTCACTACCAACGCGTTCATCGGTTTCGCGGCGGACTTGGACCCCAAC 1385
DB 1429 AGCGGCTGCTCATCTACCAACGCGTTCATTTGCGTTTTCGCGGACTTGGACCCCAAC 1488
QY 1386 ACCGCGGTTGTTGGTGAACCTGCGCAAGTACCAAGTACCAAGCAGC-----CAGGGCAACAA 1439
DB 1489 ACCGCGGTTGTTGGTGAAGTGGCGGAGTACCAAGTACCAAGCAGC-----CAGGGCAACAA 1548
QY 1440 TTGATGATGATCAACGCTTGGGCTTGTACCGGCAAGGACAACTTCGCGACCGCTGC 1499
DB 1549 TTGATGATGATCAACGCTTGGGCTTGTACCGGCAAGGACAACTTCGCGACCGCTGC 1608
QY 1500 TACGAGCGTGTGATGACCAACCGC 1523
DB 1609 TACGAGCGTGTGTTCTTCAACCGC 1632

RESULT 2

US-08-379-926A-7
; Sequence 7, Application US/08379926A
; Patent No. 5783414
; GENERAL INFORMATION:
; APPLICANT: CARREZ, DIRK
; APPLICANT: ROOS, JOEL
; TITLE OF INVENTION: EXPRESSION SYSTEM, INTEGRATION
; TITLE OF INVENTION: VECTOR
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS: OBLON, SPIVAK, MCCLELLAND, MAIER &
; ADDRESS: NEUSTADT
; STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/379,926A
; FILING DATE: 27-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: BE 09400102
; FILING DATE: 28-JAN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: BE 09400586

TRANSFORMED BY THIS INTEGRATION VECTOR

/	FILING DATE:	17-JUN-1994
/	PRIOR APPLICATION DATA:	
/	APPLICATION NUMBER:	BE 09500014
/	FILING DATE:	09-JAN-1995
/	ATTORNEY/AGENT INFORMATION:	
/	NAME:	OBLON, NORMAN F.
/	REGISTRATION NUMBER:	24,618
/	REFERENCE/DOCKET NUMBER:	3987-13-0
/	TELECOMMUNICATION INFORMATION:	
/	TELEPHONE:	703-413-3000
/	TELEFAX:	703-413-2220
/	TELEX:	248955 OPAT UR
/	INFORMATION FOR SEQ ID NO.: 7:	
/	SEQUENCE CHARACTERISTICS:	
/	LENGTH:	2045 base pairs
/	TYPE:	nucleic acid
/	STRANDEDNESS:	single
/	TOPOLOGY:	linear
/	MOLECULE TYPE:	DNA (genomic)
/	US-08-379-926A-7	
Query Match 11.3%; Score 173; DB 1; Length 2045;		
Best Local Similarity 56.6%; Pred. No. 4.5e-33;		
Matches 426; Conservative 0; Mismatches 310; Indels 17; Gaps 5;		
QY	274	GAGGACTGCCTCACCATCAACGTGGTGGCGCGCGGCCACCAAGGGCGGCGCCAACTC 333
Db	70	GAGGATTGGCTGAACATTTGACATTCGGCGTCGGCGCGGGACCACCGGGGACTCGAAGCTG 129
QY	334	CCGGTCATGCTCGATCTTTGGCGGTGGTTTTGAGATCGGCAGCCCCACCATCTTCCT 393
Db	130	CCTGTGCTGGTCTGGATCTTTGGCGGAGGCTTTGAAC TTGGTTCAAAGCGATGTATGAT 189
QY	394	CCC GCCCAGATGGTCACCAAGATGTGCTCATGGGCAAGCC-ATCATCCAAGTCGCCGTC 452
Db	190	GGTAACAAGATGGTATCATCGTCGTAGACAAGAACATGCTATCTGTTTGTTAGCAATG 249
QY	453	AATACCGTGTGCTCGTGGGGTCTTTGGCTGGATGATCATCAAGGCCGAGGGCAGC 512
Db	250	AATTATCGCTGGGAGGTTTCGGGTCTTGCCCGAAAAGGAGATCTTGGAGGACGGGTCC 309
QY	513	GGGAACCCGCGCTTGAAGGACCCAGCTTTGGGCATGCAAGTGGTGGCAGACAACTTGCC 572
Db	310	CGGAACCTAGGGCTCTTGGAACCAACGCTTGCCCTGCAGTGGGTGTCGCGAACATCGAG 369
QY	573	GGGTTCCGCGCGACCCGAGCAAGGTGACATCTTTGGCGAGGGCGGCGAGCATGTCGTGT 632
Db	370	GCCTTTGCTGGAGACCCCGGACAAGTGCAGATTGGGGAGAATCAGCAGGA-GCCATTGC 428
QY	633	TGTGCCACTCATCTGGAAACGACCGCGACAAACGTAACAAGGGCAAGCCGTTGTTCCGCG 692
Db	429	TTTGACTAGATGACTTGTAACGACGAAACATCACTTACAAGATAAGCCCTTGTTCCGGG 488
QY	693	CGGCGCATATGCGAGGAGCCATGTGC-----CGGACCCGCTGGAGCGGACGTAACGGCA 746
Db	489	GGGCGCATATGAGCTCCGCTAGTGTGTTCCCGCAGACCCCGTCGATGGGGTCAAGGGAC 548
QY	747	ACGAGATCTACGACCTCTTTGTCTCGAGTGTGGCTGTGGCAGCGCCAGCGACAAGCTCG 806
Db	549	AGCAAGTATATGATCGGTAATGGAATCTCGAGGCTGTTCTCTTTAAAGCAGACCCCTAG 608
QY	807	CGTGCTTGGCAGTGCAGGACAC----CTTGCTCGATGCCACCAACAACTCTCTGGGT 863
Db	609	CTTGCTCGCTGAAC TACAGCTACACCGACTTCTCAATGCGGCAAACTCCGTGCCAGGCA 668
QY	864	TCITGGGGTACTCCTCGTTGGGTTGTA CT-----CCGCGCGGACGGGCAAGAACATCA 917
Db	669	TTTTAAGCTACCAATCTGTGGCGTTATCATATGTGCTCGACCGGACGGGACGGCGTGT 728
QY	918	CCGATGACATGTACAAGTTGCTGGCGGACGGCAAGTATGCAAGCGTTTCCCGTGCATCTG 977
Db	729	CGGCATCACCGGACGTTTGGGCAAGCAGGGAATATGCTCGGGTCCCCTTCATCGTGG 788

US-08-362-525-11

Query Match 10.7%; Score 164.2; DB 3; Length 1828;
Best Local Similarity 54.6%; Pred. No. 6.3e-31;
Matches 436; Conservative 0; Mismatches 328; Indels 35; Gaps 4;
QY 248 TGTTCAGGCGGTCTTCCACAGTGTGAGTGTGCTTCCATCAAGTGTGCGCGC 307
DB 377 TGGCCAGGAGTGTCTCCATGATGAGGACTGTCTACTTAAAGTTCGCGCCG 436
QY 308 CGGCACCAAGCGGCGCAACCTCCCGGTCTGATCTGTGATCTTTGGGGTGGTTG 367
DB 437 CTGGACCAAGCCTGATGCTAAGCTCCCGGTCTGATGTTGATTTACGGTGTGCTTTG 496
QY 368 AGATCGGAGCGCCACCATCTTCCCTCCCGCCAGATGTGTACCAAGTGTGTCTATGG 427
DB 497 TGTGTGTTCTTCTGTCTTACCTGTGTACCGCTAGCTCAAGGAGAGTGTGAAATGG 556
QY 428 GCAGCC-ATCATCCAGTGTGCGCTCAACTACCGTGTGTGCTGTGGGGTCTTTGGCTG 486
DB 557 GCCAGCCTGTGTGTTGTTTCCATCACTACCGTACCGGCCCTATGATCTTGGTG 616
QY 487 GTGATGATCAAGCGCGGAGCGGAGCGCGGAAACCGCGCTTGAAGCAAGCGTTTGGGCA 546
DB 617 GTGATGCCATCACCGCTGAGGCGCAACCAACCGCTGTCTGACGACGAGCGCAAGGTC 676
QY 547 TGCAGTGTGTGAGCAACAATTGCGCGGTTCGGGGCGGACCCGAGCAAGTGA-CA 602
DB 677 TCGAGTGGGTGTAGCGACAACATTTGCGTGTGTGATCTTCCGACAAAGTGTGATTT 736
QY 603 TCTTTGGCGAGCGGCGAGCATGTCGCTGTGTGCGACCTCATCTGGAACGACGCGCA 662
DB 737 TCGGTGAGTCCGCTGTGTCCTCAAGTGTGTCTACCAAGTGTGTGCTTACGGTGTGACA 796
QY 663 ACACGTACAGGCGAAGCGGTGTTTCCGCGGCGCATCATGCAAGGAG-710
DB 797 ACACCTACAGGAAAGCAGCTTTTCCACTCTGCGATCTTCAAGTGTGCGGCTCTTTC 856
QY 711 ---CATGCTGCGGACCCGCTGGAGCGGACGTACGCGCAACGAGATCTACGACCTTTG 767
DB 857 CTTACTTTGACTCTACTTCTGTGTGCTGCGAGTGTGCTTACAGAGATTTGCTCAGTATG 916
QY 768 TCTGAGTGTGCTGTGCGAGCGGCGAGCAAGCTGCGGTGCTTGGCGAGTGGGAGG 827
DB 917 CGGATGTGACACAGTGTGCTGATGATGACACTCTGCTGTGTCTCCGCAAGTCA 976
QY 828 ACACCTTGTGATGCCACCAACAACTCTGCGGTGCTTGGCGTACTCTCTGCTG-884
DB 977 GCGATGTCTTGACAGTGTGCGAAGTCTGATGATCTTAAAGGACCTGTTGCTGCTCC 1036
QY 885 -----GGTTGACTCCCGCGCGCGGCAAGCAATCAACGATGATGTACA 932
DB 1037 CTTCAATTCCTTGGATTTGGTCCGACCGCGCAAGCAATTTATTCGATGCGCTTATG 1096
QY 933 AGTTGCTGCGGACGCGCAAGTATGCAAGCGTTTCCCGTGTATCATTTGCGCAACGAGCG 992
DB 1097 AGCTTACCGGAGCGGTAGATACGCAAGTTCCCTACATTAAGTGTGCAACGAGGAGTG 1156
QY 993 AGGCAACCATCTTTGGCTC 1011
DB 1157 AGGCTACTATTCTTGCCCC 1175

RESULT 4
US-08-776-210-5
; Sequence 5, Application US/08776210
; Patent No. 5942659
; GENERAL INFORMATION:
; APPLICANT: ALIBERT, Gilbert
; APPLICANT: MOULOUNGUI, Zephirin
; APPLICANT: BOUDET, Alain
; TITLE OF INVENTION: PROCESS FOR PRODUCING FATTY ACIDS OR
; TITLE OF INVENTION: DERIVATIVES THEREOF FROM OLEAGINOUS PLANTS

NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: YOUNG & THOMPSON
STREET: 745 South 23rd Street
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/776,210
FILING DATE: 24-JAN-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 94 09272
FILING DATE: 25-JUL-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR95/00957
FILING DATE: 18-JUL-1995
ATTORNEY/AGENT INFORMATION:
NAME: PATCH, Andrew J
REGISTRATION NUMBER: 32,925
REFERENCE/DOCKET NUMBER: IN 387 - BE 6996
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-521-2297
TELEFAX: 703-685-0573
TELEX: 248425 EMBON
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1635 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
US-08-776-210-5
Query Match 10.6%; Score 161.8; DB 2; Length 1635;
Best Local Similarity 55.9%; Pred. No. 2.3e-30;
Matches 447; Conservative 0; Mismatches 317; Indels 35; Gaps 6;
QY 248 TGTTCAGGCGGTGCTTCCACAGTGTGAGTGTGCTTCCATCAAGTGTGCGCGC 307
DB 281 TGGCCAGGAGTGTCTCTCCATGATGAGGACTGTCTACTTAAAGTTCGCGCCTG 340
QY 308 CGGCACCAAGCGGCGGCGCAACCTCCCGGTCTGATCTTGGATCTTTGGCGGTGGTTG 367
DB 341 CTGGACCAAGCCTGATGCTAAGCTCCCGCTCATGTTGGATTTACGTTGTCCTTTG 400
QY 368 AGATCGGACGCGCCACCATCTTCCCTCCCGCCAGATGTGTCAACGAGTGTGCTCATGG 427
DB 401 TGTGTTGTTCTTCTGTTCTTACCTCTGTTAAACGCTACGTAAGGAGAGTGTGAAATGG 460
QY 428 GCAAGCC-ATCATCAAGTGTGCGCGTCAACTACCGTGTGCTTCCCTGCGGGTCTTTGGCTG 486
DB 461 GCCAGCCTGTTGTTGTTTCCATCACTACCGTACCGGCCCTATGGAATTCCTGGGTG 520
QY 487 GTGATGATCAAGCGCGGAGCGGAAACCGCGCTTGAAGGACCAAGCGTTTGGGCA 546
DB 521 GTGATGCCATCACCGTGTGAGGTGTACCAACGCTGTGTGACGACCAAGCGCAAGGTC 580
QY 547 TGCAGTGTGTCGACACAACTTTCGCGGTTCGCGGCGGACCGAGCAAGTGA-CA 602
DB 581 TCGAGTGGGTAGCGACAACTTTCGCAACTTTGTTGTTGATCCCGCAAGGTCATGATTT 640
QY 603 TCTTTGGCGAGGCGGCGACCATGTGCTGTTGTGCGACCTCATCTCTGGAAGCGGCA 662
DB 641 TCGGTGAGTCCGCTGTGTCATGAGTGTGCTACCAAGTGTGTTGCTTACGTTGGTGTGACA 700
QY 663 ACACGTACAGGCGCAAGCGCTTGTTCGCGCGCGGACATGTCAGGAGGAG-----710

Db 701 ACACCTACACGGAAGAGCTTTTCCACTCTGCCATTCTTCAGTCTGGCGGTCTCTTC 760
QY 711 ---CCATGTCGCGGACCCGCTGACGGCAGTACGACGATCTAGACCTCTTTG 767
Db 761 CTATCTTGTACTTCTTCTTGTGTCGAGAGTCTACAGAGATTTGCTCAGTATG 820
QY 768 TCTCGAGTGTGCTGTGTCGAGCGCCAGCAGCAGTCTGCTGCTGCTGCTGCTG 824
Db 821 CCGGATGTGATGTCAGCGCCAGTGACAACTGAACTCTGCTGCTGCTGCTGCTG 880
QY 825 GCACACACTTGTCTGATGTCACCAAC-----AACACTCTGCTGCTGCTGCTGCTG 877
Db 881 GCATGTCTTGTGACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 940
QY 878 TCGTTGCGGTGTGACT-----CCGCGCCGACGAGTCTGCTGCTGCTGCTGCTGCTG 932
Db 941 CTCAAATCTTGTGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1000
QY 933 AGTTGTCGCGGACGCGGACGATGTCAGAGCTTCCGCTGATCAITGGCGACGAGACG 992
Db 1001 AGCTTACCGCAGCGGTAGATACGCCAAGTTCCTTACATTAATCTGTTAACGAGGATG 1060
QY 993 AGGGACCACTTTTGGCTC 1011
Db 1061 AGGGTACTATTCTTGGCCC 1079

RESULT 5

US-08-932-376A-1
; Sequence 1, Application US/08932376A
; Patent No. 5869309
; GENERAL INFORMATION:
; APPLICANT: Politino, Michael
; APPLICANT: Tonzi, Sean M.
; APPLICANT: Usher, John J.
; APPLICANT: Burnett K, William V.
; APPLICANT: Romangik, Guna
; TITLE OF INVENTION: CEPHALOSPORIN ESTERASE GENE FROM
; TITLE OF INVENTION: RHODOSPORIDIUM TORULOIDES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bristol-Myers Squibb Company
; STREET: Rt. 206 & Provinceline Road
; CITY: Princeton
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 08543-4000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/932,376A
; FILING DATE: 17-SEP-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Savitsky, Thomas R.
; REGISTRATION NUMBER: 31,661
; REFERENCE/DOCKET NUMBER: ON0144a
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (609) 252-4956
; TELEFAX: (609) 252-4526
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1738 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS

; LOCATION: 11..1726
US-08-932-376A-1

Query Match 5.8%; Score 88.2; DB 2; Length 1738;
Best Local Similarity 55.2%; Pred. No. 2,7e-12;
Matches 218; Conservative 0; Mismatches 168; Indels 9; Gaps 2;
QY 274 GAGGACTCGCTCACCACCAAGCTGTGCGCGCCGCGGACCAAGCGGCGGCGCAACCTC 333
Db 371 GAGGATTCCTCTCTCTCAATGCTGTTGCCCGCGGCTCTGTCGAGGCGGCAATCTT 430
QY 334 CCGGTCTGCTGTGATCTTTTGGCGGTGGTGTGAGATCGGCGAGCCGCCACCATCTTCCCT 393
Db 431 CCGGTCTGCTGTGATCTTTTGGCGGTGGTGTGAGATCGGCGAGCCGCCACCATCTTCCCT 393
QY 394 CCGCGCCAGATGTGTCACCAAGAGTGTGCTCATGGGCAAGCCATCATCCAGTGGCGTCA 453
Db 491 GACTTTGCCGCTTTCACCAAGCACAG-----GGAACCAAGATGGTGTGTAATCTCC 545
QY 454 ACTACCGTGTGCTGCTGCGGGGTCTTGGCTGCTGATGACATCAAGGCGGCGGAGCG 513
Db 546 AGTACCGTCTGCGAGCTTTGGTTTCTGCTGCGCAAGCCATGAAGACTACGGGTAA 605
QY 514 GGAACCGCGCTTTGAAGGACCAAGCTTTGGGCAATGCAATGGGTGGGAGACAACATTGCCG 573
Db 606 CGAACCGCGCTTTGCTTGACCAAGCAATTCGCCCTTCAATGGGTTCACAGCACGTCTCGA 665
QY 574 GGTTCGCGCGGACCCGAGCAAGTGTGATCTTTGG-----CGAGCGGCGGAGCATGTCCG 629
Db 666 AGTTGCGCGGCAACCCCGATCAGTTTACGATTTGGGCGAGTCTGAGGCGGCGGTTCCG 725
QY 630 TGTGTGCCACCTCATCTGGAACGACGCGGACAAAC 664
Db 726 TTAAGAACAGATCATTTGGAACGCGGCGGCAACACC 760

RESULT 6

US-07-846-181-6
; Sequence 6, Application US/07846181
; Patent No. 5360732
; GENERAL INFORMATION:
; APPLICANT: BERKA, RANDY M
; APPLICANT: FOWLER, TIMOTHY
; APPLICANT: REV. MICHAEL W
; TITLE OF INVENTION: PRODUCTION OF ASPERGILLUS NIGER
; TITLE OF INVENTION: CATALASE-R
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENENCOR INTERNATIONAL, INC.
; STREET: 180 KIMBALL WAY
; CITY: SOUTH SAN FRANCISCO
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/846,181
; FILING DATE: 19920304
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: HORN Ms, MARGARET A
; REGISTRATION NUMBER: 33401
; REFERENCE/DOCKET NUMBER: GC204-US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-742-7536
; TELEFAX: 415-742-7217
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8533 base pairs

TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-07-846-181-6

Query Match 5.7%; Score 87; DB 1; Length 8533;

Best Local Similarity 50.8%; Pred. No. 9.8e-12; Indels 16; Gaps 7;
Matches 389; Conservative 0; Mismatches 360;

274 GAGGACTGCTCACCACCAAGTGTGCTCATGGGCAAGCC-ATCATCCAGTGGCCGTC 452
Db 12 GAGGATTGCTGAACATTTGACATTCGGCGCCAGCAACCCAACTGGCGAGCGGAATGC 71
Qy 334 CCGGTCTATGCTCTGATCTTTGGCGGTGGTTGAGATCGGAGCCCAACCATCTCCCT 393
Db 72 CCGGTCTGCTGCTGATCTTTGGCGGAGGCTTGAACCTGGTTCAAAGGCCATGTATGAC 131
Qy 394 CCGGCCACAGATGGTCAACCAAGTGTGCTCATGGGCAAGCC-ATCATCCAGTGGCCGTC 452
Db 132 GGCACAAAGATGGTATCATGCTCGATAGCAAGATATGCTATCGTTTGTAGCGATG 191
Qy 453 AACTACCGTGTGCTCGTGGGGTCTTGGCTGATGATCATCAAGCGCGGAGGCGAGC 512
Db 192 AACTATCGGTAGGGGCTTCGGGTTTCTGCCGGAAGAGGATTTCTGGAGGACGGGTCC 251
Qy 513 GGGAAACCGCGTGAAGGACCGAGGTTTGGGATGCGAGTGGTGGGCAAGCAATTCGCC 572
Db 252 GCCAACTTAGG--TCTTTGACCAAGCCTTGCCCT---AGTGGGTGGCGCAACATCGAG 306
Qy 573 GGGTTCGGCGGCGCCAGCAAGTGTGCTCATGGGCAAGCC-ATCATCCAGTGGCCGTC 631
Db 307 GGGTTTGTGGAGACCAAGAGTGTGACAACTGGGGAGATCAGCAGGGGCTATTTCTG 366
Qy 632 TTGTGCCACCTCATCTGGAACGACGCGGCAACAGTCAAGGCGGAGCGGTGTTCGCG 691
Db 367 TCTTGATCAGATGATCTGTACGACGGAACATCGCTTACAGGCAAGCCCTTGTTCGG 426
Qy 692 GCGG----GCATCATGCAAGGACCATGTTGCCGACCCGGTGGAGCGCAGTACGGCAA 747
Db 427 GAGCCATCATGAGCTCCGGTATGTGTTCCGAGACCCCTGTCCGAGGGGTCAAGGATCA 486
Qy 748 CGAGATCTACGACCTCTTTGTCTCGAGTGTCTGGTGGAGCGGCGGAGCAAGTTCGC 807
Db 487 GCAAGTATGATCGGTGTGGAGCTCTGAGGCTGTTCCTCTTCCACGACACCTTGC 546
Qy 808 GTGCTTGGCGAGTGGAGCGACAC---CTTGCTCGATGCGCAACAAACACTCTCGGTT 864
Db 547 TTGTCTGCTGAGCTAGACTACACGACTATCTCAATGCGGCAAACTCGTCCGGGATC 606
Qy 865 CTGCGGTACTCTCTGCTGGGTGTGTTCTCCGCGCCGACCGGCAAGACATCACCGATGA 924
Db 607 CTAGGTATACCGTGGGCTATCATATGTGCTCGACAGAGGAGCGGCAATTTGTGCG 666
Qy 925 CAT--GTACAAGTTGGTGGCGAGCAAGTATGCAAGGTTTCCGTTGATCATTTGCGCAG 982
Db 667 GGTGCGCAATTTGGGTAAGCAGGGAATGATGCGGGTCCCAATTCATGCTGGGCGAC 726
Qy 983 CAGAACGACGAGGCAACATCTTTGGCTCTTGAACGTGACCGA 1027
Db 727 CAAGAGGATGAGGGGACCTTGTTCGCTTGTTCAGTCTCCTTACGA 771

RESULT 7

US-07-845-989-6
Sequence 6, Application US/07845989
Patent No. 5360901
GENERAL INFORMATION:
APPLICANT: BERKA, RANDY M
APPLICANT: FOWLER, TIMOTHY
APPLICANT: REY, MICHAEL W
TITLE OF INVENTION: PRODUCTION OF ASPERGILLUS NIGER
TITLE OF INVENTION: CATALASE-R

NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENENCOR INTERNATIONAL, INC.
STREET: 180 KIMBALL WAY
CITY: SOUTH SAN FRANCISCO
STATE: CA
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/845,989
FILING DATE: 19920304
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: HORN MS, MARGARET A
REGISTRATION NUMBER: 33401
REFERENCE/DOCKET NUMBER: GC208-US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-742-7536
TELEFAX: 415-742-7217
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 8533 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-07-845-989-6

Query Match 5.7%; Score 87; DB 1; Length 8533;
Best Local Similarity 50.8%; Pred. No. 9.8e-12; Indels 16; Gaps 7;
Matches 389; Conservative 0; Mismatches 360;

Qy 274 GAGGACTGCTCACCACCAAGTGTGCTCATGGGCAAGCC-ATCATCCAGTGGCCGTC 393
Db 12 GAGGATTGCTGAACATTTGACATTCGGCGCCAGCAACCCAACTGGCGAGCGGAATGC 71
Qy 334 CCGGTCTATGCTCTGATCTTTGGCGGTGGTTGAGATCGGAGCGGCGGCAACCTC 393
Db 72 CCGGTCTGCTGCGATCTTTGGCGGAGGCTTTGAACTTGGTTCAAAGGCCATGTATGAC 131
Qy 394 CCGGCCACAGATGGTCAACCAAGTGTGCTCATGGGCAAGCC-ATCATCCAGTGGCCGTC 452
Db 132 GGCACAAAGTGTATCATGCTCGATAGCAAGATATGCTATCGTTTGTAGCGATG 191
Qy 453 AACTACCGTGTGCTCGTGGGGTCTTTGGCTGATGATCATCAAGCGCGGAGGCGAGC 512
Db 192 AACTATCGGTAGGGGCTTCGGGTTTCTGCCGGAAGAGGATTTCTGGAGGACGGGTCC 251
Qy 513 GGGAAACCGCGTGAAGGACCGAGGTTTGGGATGCGAGTGGTGGGCAAGCAATTCGCC 572
Db 252 GCCAACTTAGG--TCTTTGACCAAGCCTTGCCCT---AGTGGGTGGCGCAACATCGAG 306
Qy 573 GGGTTCGGCGGCGCCAGCAAGTGTGCTCATGGGCAAGCC-ATCATCCAGTGGCCGTC 631
Db 307 GGGTTTGTGGAGACCAAGAGTGTGACAACTGGGGAGATCAGCAGGGGCTATTTCTG 366
Qy 632 TTGTGCCACCTCATCTGGAACGACGCGGCAACAGTCAAGGCGGAGCGGAGCAAGTTCGC 807
Db 487 GCAAGTATGATCGGTGTGGAGCTCTGAGGCTGTTCCTCTTCCACGACACCTTGC 546

Qy	808	GTGCTTGGCAGTGGAGCGACAC--CTTGCTGATGCCACCAACACACTCTCTGGGTT	864
Db	547	TTGTCTGCTGAGCTAGACTACCCGACTATCTCAATGCGCAACTCTGTGCGGGATC	606
Qy	865	CTTGGCGTACTCTCTGTTGCGTGTACTCCGGGCCGACGCAAGACATCACCGATGA	924
Db	607	CTAGGTTATCACCGTGGCGCTATCATATGTGCTCGACGACGGACGGCAATTGTTCGG	666
Qy	925	CAT--GTACAAGTTGGTSCGCGACGCGCAAGTATGCAAGCTTCCCGTGATCATTTGGCGAC	982
Db	667	CGTCGCCAGAATTTTGGGTAAAGCAGGGAAGTATGCGGGTCCCAATTATCGTGGCGGNC	726
Qy	983	CAGAACACGAGGGCACCATCTTTGGCTCTTGAACGTGACCAAGA	1027
Db	727	CAAGAGATGAGGGACCTGTCTGCCCTGTGTTTCACTGCTCTTACGA	771

RESULT 8

```

US-08-445-050-8
; Sequence 8, Application US/08445050
; Patent No. 5763739
; GENERAL INFORMATION:
; APPLICANT: Blaeckberg, Lars
; APPLICANT: Edlund, Michael
; APPLICANT: Hansson, Lennart
; APPLICANT: Hernell, Olle
; APPLICANT: Lundberg, Lennart
; APPLICANT: Stromqvist, Mats
; APPLICANT: Toernell, Jan
; TITLE OF INVENTION: No. 5763739el Polypeptides
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/445,050
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/204,691
; FILING DATE:
; APPLICATION NUMBER: SE 9300686-4
; FILING DATE: 01-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 9300722-7
; FILING DATE: 04-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner Ph.D., Richard J
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-850
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)819-8783
; TELEFAX: (212)354-8113
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2184 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens

```

```

RESULT 9
US-08-204-691-8
; Sequence 8, Application US/08204691
; Patent No. 5827683
; GENERAL INFORMATION:
; APPLICANT: Blaeckberg, Lars
; APPLICANT: Eklund, Michael
; APPLICANT: Hansson, Lennart
; APPLICANT: Herneli, Olle
; APPLICANT: Lundberg, Lennart
; APPLICANT: Stromqvist, Mats
; APPLICANT: Toernell, Jan
; TITLE OF INVENTION: No. 5827683el Polypeptides
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/204,691
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 9300686-4
; FILING DATE: 01-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 9300722-7
; FILING DATE: 04-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner Ph.D., Richard J
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-850
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)819-8783
; TELEFAX: (212)354-8113
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2184 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: mammary gland
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 82..2088
; OTHER INFORMATION: /label= Variant_T
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 151..2085
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 1756..2052
; FEATURE:
; NAME/KEY: repeat_unit
; LOCATION: 1756..1788
; FEATURE:
; NAME/KEY: repeat_unit
; LOCATION: 1789..1821
; FEATURE:
; NAME/KEY: repeat_unit
;

```


Db 698 GGGACCCCAACATCACGCTCTTCGGGGAGTCTGCTGGAGGTGC 743

RESULT 11
US-08-204-691-1
Sequence 1, Application US/08204691
Patent No. 5827683
GENERAL INFORMATION:
APPLICANT: Blaeckberg, Lars
APPLICANT: Edlund, Michael
APPLICANT: Hansson, Lennart
APPLICANT: Hernel, Olle
APPLICANT: Lundberg, Lennart
APPLICANT: Stroemqvist, Mats
APPLICANT: Toernell, Jan
TITLE OF INVENTION: NO. 5827683el Polypeptides
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: White & Case
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10036-2787
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/204,691
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: SE 9300686-4
FILING DATE: 01-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: SE 9300722-7
FILING DATE: 04-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Sterner Ph.D., Richard J
REGISTRATION NUMBER: 35,372
REFERENCE/DOCKET NUMBER: 1103326-850
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)819-8783
TELEFAX: (212)354-8113
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2428 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: mammary gland
FEATURE:
NAME/KEY: CDS
LOCATION: 82..2319
OTHER INFORMATION: /product= "bile-salt-stimulated
OTHER INFORMATION: lipase"
FEATURE:
NAME/KEY: exon
LOCATION: 985..1173
FEATURE:
NAME/KEY: exon
LOCATION: 1174..1377
FEATURE:
NAME/KEY: exon
LOCATION: 1378..1575

FEATURE:
NAME/KEY: exon
LOCATION: 1576..2415
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 151..2316
FEATURE:
NAME/KEY: polyA_signal
LOCATION: 2397..2402
FEATURE:
NAME/KEY: repeat_region
LOCATION: 1756..2283
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..81
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 1756..1788
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 1789..1821
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 1822..1854
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 1855..1887
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 1888..1920
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 1921..1953
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 1954..1986
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 1987..2019
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 2020..2052
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 2053..2085
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 2086..2118
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 2119..2151
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 2152..2184
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 2185..2217
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 2218..2250
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 2251..2283
US-08-204-691-1

Query Match 4.0%; Score 61.2; DB 1; Length 2428;
Best Local Similarity 52.7%; Pred. No. 1.3e-05;
Matches 214; Conservative 0; Mismatches 173; Indels 19; Gaps 3;
QY 236 TGCAGTCCAAAGGTGTTCCAGGGGGTTCCTCCAGAGTGGAGTGCCTTCACCATCAACG 295
DB 344 TGCAGGCCACCATCACCAGGAGACACACCTACGGGGATGAGAGCTGCCTGTACCTCAACA 403

```
Qy 296 TGGTGGCGCCCGCGGCGACCAAGCGCGCGCC---AACTCCCGGTATGCTCTGGATCT 352
Db 404 TTTGGGTGCCCGCCAGGCGAGGAAGTCTCCCGGACCTGCCCGTTATGATCTGGATCT 463
Qy 353 TTGGCGGTGGGTTTGAGATCGGCAGC-----CCACCATCTTCCCTCCCGCCAG 402
Db 464 ATGAGGCGCCTTCTCATGGGGTCCGGCCCATNGGGCCAACTTCTCAACAACCTACCTGT 523
Qy 403 ATGCTCACCAAGAGTGTCTCATGGCAAGCATCATCCAGTGGCGGTCACTACCGTG 462
Db 524 ATGACGGCGAGGATCGCCACACGCGGAACGTCATCGTGTCACTTCACTTACCGTG 583
Qy 463 TTGCTCTGTGGGGTCTTGTGCTGGTGAATCAATCAAGCGCGGAGCGGGAACCGCG 522
Db 584 TCGCCCGCCTTGGGTTCTCAGCACTGGGAC-----GCCAATCTGCCAGGTAATG 637
Qy 523 GCTTGAAGGACCGCGTTTGGGCATGAGTGGGTGGGTCAGACAACTTCCCGGTTCCGGG 582
Db 638 GCCTTCGGGATCAGCACATGGCCATGCTTGGGTGAAGAGGAATATCGCGCCTTCGGGG 697
Qy 583 GCGACCCGAGCAAGGTGACATCTTTGCGAGCGCGGCGAGCATGTCC 628
Db 698 GGGACCCCAACACATCACGCTCTTCGGGGAGTCTGCTGGAGGTGC 743
```

RESULT 12

```
US-09-355-295B-2
; Sequence 2, Application US/09355295B
; Patent No. 6525241
; GENERAL INFORMATION:
; APPLICANT: Dalrymple, M.
; APPLICANT: Lundberg, L.
; APPLICANT: Stromqvist, M.
; TITLE OF INVENTION: Expression Methods
; FILE REFERENCE: 1754 SEQUENCE LISTING V2A.txt
; CURRENT APPLICATION NUMBER: US/09355,295B
; PRIOR FILING DATE: 1999-07-18
; PRIOR APPLICATION NUMBER: PCT/SE99/00648
; PRIOR FILING DATE: 1999-04-21
; NUMBER OF SEQ ID NOS: 5
; SEQ ID NO 2
; LENGTH: 2428
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: cDNA
; FEATURE:
; OTHER INFORMATION: Mammary gland source
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 82..2319
; OTHER INFORMATION: /product= "bile salt-stimulated
; OTHER INFORMATION: lipase"
; FEATURE:
; NAME/KEY: exon
; LOCATION: 985..1173
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1174..1377
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1378..1575
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1576..2415
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 151..2316
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 2397..2402
; FEATURE:
; NAME/KEY: repeat_region
```

Query Match 4.0%; Score 61.2; DB 4; Length 2428;

Best Local Similarity 52.7%; Pred. No. 1.3e-05;

Matches 214; Conservative 0; Mismatches 173; Indels 19; Gaps 3;

Qy 236 TGCAATCCAAAGGTGTTCCAGGGCGGTGCTTCCCGAGAGTGAGGACTGCCTCACCATCAACG 295

Db 344 TGCAGGCCACCATCACCCAGGACAGCACCTACGGGATGAAGACTGCCTGTACCTCAACA 403

Qy 296 TGGTGGCGCGCGCGGCGACCAAGCGCGCGCC---AACTCCCGGTATGCTCTGGATCT 352

Db 404 TTTGGGTGCCCGCCAGGCGAGGAGTCTCCCGGACCTGCCCGTTATGATCTGGATCT 463

Qy 353 TTGGCGGTGGGTTTGAGATCGGCAGC-----CCACCATCTTCCCTCCCGCCAG 402

Db 464 ATGAGGCGCCTTCTCATGGGGTCCGGCCATGGGGCCAACTTCTCAACAACCTACCTGT 523

Qy 403 ATGGTCAACAAGAGTGTGCTCATGGGCAAGCCATCATCCAGTGGCGGTCAACTACCGTG 462

Db 524 ATGACGGCGAGGAGATCGCCACACGCGGAACGTATCATCGTGGTCACTTCAACTACCGTG 583

426 TTTGGGTGCCCCAGGGCAGGAAGCAAGTCTCCCGGGACCTCCCGCTTATGATCTGGATCT 485
 353 TTGGCGGTGGGTTTGAGATCGGCAGC-----CCCAACCATCTTCCCTCCCGCCCGAG 402
 486 ATGAGAGCGGCTTCTCTCATGGGTCCGGCCATGGGGCAACTCTCTCAACAACATACCTGT 545
 403 ATGTGTCAACAGAGTGTGCTCATGGGCAAGCCATCTCCAGTGGCGGTCACTACCTG 462
 546 ATGACGGCGAGGAGATCGCCACACGCGGAAACGTCATCTGTGTGTCACCTTCAACTACCGTG 605
 463 TTGCTCGTGGGGGTTCTTGGCTGTGATGATCAAGGCGGAGCGGAGCGGAACGCGG 522
 606 TCGGCCCCCTTGGGTTCTCAGCACTGGGGAC-----GCCAATCTGCCAGGTACTATG 659
 523 GCTTGAAGGACCGAGCTTTGGGCATGTCAGTGGGTGGCAGACAAATTTGCCGGTTCGGCG 582
 660 GCCTTCGGGATCAGCACATGCCATTGCTTGGGTGAAGAGGAATATCGCGGCTTCGGGG 719
 583 GCGACCGGAGCAAGGTGACATCTTTGGGAGCGGCGGAGCATGTCC 628
 720 GGGACCCCAACAACATCACGCTCTTCGGGGAGTCTGCTGGAGGTGC 765

RESULT 14

US-07-732-962A-1
 ; Sequence 1, Application US/07732962A
 ; Patent No. 5248604
 ; GENERAL INFORMATION:
 ; APPLICANT: Fischer, Meir
 ; TITLE OF INVENTION: EXPRESSION OF ENZYMATICALLY ACTIVE
 ; NUMBER OF SEQUENCES: 2
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: John P. White, Esq.
 ; STREET: 30 Rockefeller Plaza
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10112
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07732,962A
 ; FILING DATE: 19910722
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: White, John P.
 ; REGISTRATION NUMBER: 28,678
 ; REFERENCE/DOCKET NUMBER: 39304/JPW/LSW
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 977-9550
 ; TELEFAX: (212) 664-0525
 ; TELEX: 422523 COOP UI
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1845 base pairs
 ; TYPE: NUCLEIC ACID
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 1..1842
 ; US-07-732-962A-1

Query Match 3.9%; Score 60; DB 1; Length 1845;
 Best Local Similarity 53.1%; Pred. No. 2.3e-05;
 Matches 203; Conservative 0; Mismatches 165; Indels 14; Gaps 3;

463 TTGCTCGTGGGGTCTTGGCTGGTGTGATGACATCAAGCGCGAGCGGAGCGGAACGCG 522
 584 TCGGCCCCCTTGGGTTCTCAGCACTGGGGAC-----GCCAATCTGCCAGGTACTATG 637
 523 GCTTGAAGGACCGAGCTTTGGGCATGTCAGTGGGTGGCAGACAAATTTGCCGGTTCGGCG 582
 638 GCCTTCGGGATCAGCACATGCCATTGCTTGGGTGAAGAGGAATATCGCGGCTTCGGGG 697
 583 GCGACCGGAGCAAGGTGACATCTTTGGGAGCGGCGGAGCATGTCC 628
 698 GGGACCCCAACAACATCACGCTCTTCGGGGAGTCTGCTGGAGGTGC 743

RESULT 13

US-08-370-223-12
 ; Sequence 12, Application US/08370223
 ; Patent No. 6107026
 ; GENERAL INFORMATION:
 ; APPLICANT: Lange, III, Louis G.
 ; APPLICANT: Kumar, B. Vijaya
 ; TITLE OF INVENTION: Methods and Reagents for RFLP Analysis
 ; NUMBER OF SEQUENCES: 13
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Allegretti & Witcoff, Ltd.
 ; STREET: 10 South Wacker, Suite 3000
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: USA
 ; ZIP: 60606
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/370,223
 ; FILING DATE:
 ; CLASSIFICATION: 307
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/053,308
 ; FILING DATE:
 ; APPLICATION NUMBER: 07/730,204
 ; FILING DATE: July 15, 1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: McDonnell, John J.
 ; REGISTRATION NUMBER: 26,949
 ; REFERENCE/DOCKET NUMBER: 91,441
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 312-715-1000
 ; TELEFAX: 312-715-1234
 ; INFORMATION FOR SEQ ID NO: 12:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2487 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cdna
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 104..2341
 ; US-08-370-223-12

Query Match 4.0%; Score 61.2; DB 3; Length 2487;
 Best Local Similarity 52.7%; Pred. No. 1.3e-05;
 Matches 214; Conservative 0; Mismatches 173; Indels 19; Gaps 3;
 236 TGCAGTCCAGGTGTTCCAGCGGTGCTTCCCGAGAGTGGAGTGCCTCACCATCAACG 295
 366 TGCAGGCCACCATCACCAGAGACGACCTACGGGATGAGACTGCTGTACTCTACCA 425
 296 TGGTGGCGCGCGCGGCGCACCAAGCGGGCGCC---AACCTCCCGGTGATGCTGTGATCT 352

242	Qy	CC	AAGGTGTTCC	AGCGGTGTTCC	CCAGAGTGA	AGGACTGCT	CAACATCA	ACAGTGGTGC	301
341	Db	CC	GAGATGTG	AAACCC	CAACGTTAG	CTGAGCG	AGCACTGC	TGTACCTCAAC	397
302	Qy	GG	CCGCGGG	CACCA	AGCGGG	CCCAACCT	CCCGT	CATGCTCTG	361
398	Db	GG	ACACCA	TACCCC	CGGCTT	ACATCCCC	ACCCCTGT	CCTCGTCTG	457
362	Qy	GG	TTTGAGAT	CGGAGC	CCCAACCA	TCATTCC	TCCCGCC	AGATGCTCA	421
458	Db	GC	TTCTAC	AGTGGG	CTCTCT	CTTG	AGGTG	TACGA	509
422	Qy	TC	ATGGGCA	AGC	ATCAT	CACTG	GGCGTCA	ACTACCGTGT	481
510	Db	AC	AGCCGAG	AGGACT	GTGTGT	GTCTCC	ATGA	CTACCGGTGG	569
482	Qy	GG	CTGTGT	ATGAC	ATCA	AGGCCG	AGGGCAG	CGGGNA	541
570	Db	GG	CCCTGCC	GGGAGC	---CG	AGAGGC	CCCGGCA	ATGTGGTCT	626
542	Qy	GG	GCATGC	ATGGGT	TGGCAG	ACAAT	TTCGCG	GGTTCGCG	601
627	Db	GG	CCCTGC	ATGGGT	TGCAG	GAGAC	GTGGC	AGCCTTCG	686
602	Qy	AT	CTTTGG	CAGG	CGGG	CAGCA	623		
687	Db	GC	TGTTTGG	GAGAG	CGCG	GGA	708		

RESULT 15

```

PCT-US92-06106-1
; Sequence 1, Application PC/TUS9206106
; GENERAL INFORMATION:
; APPLICANT: Fischer, Meir
; TITLE OF INVENTION: ENZYMATICALLY ACTIVE RECOMBINANT HUMAN
; TITLE OF INVENTION: ACETYLCHOLINESTERASE AND USES THEREOF
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White, Esq.
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06106
; FILING DATE: 19920722
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 39304-B-PCT/JPW/EAB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1845 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1842
PCT-US92-06106-1

```

Query Match.	3.9%;	Score 60;	DB 5;	Length 1845;
Best Local Similarity	53.1%;	Pred. No. 2.3e-05;		
Matches	203;	Conservative 0;	Mismatches 165;	Indels 14; Gaps 3;
Qy	242	CCAAGTGTTCCAGGCGTGCTTCCCCAGAGTCAGGAGTCCCTCACCATCAACGTGTGTC	301	
Db	341	CCGAGATGTGGAAACCCCAACCGTGTAGCTGAGCGAGGACTGCCTGTACTCAAC---	397	GTGT
Qy	302	GGCCGCGCGGCACCAAGCGGGGCCAACCTCCCGGTCACTCTCTGGATCTTTGGCGGTG	361	
Db	398	GGACACCATACCCCGCGCCTACATCCCCACCCCTGTCTCTCGTGTGATCTATGGGGGTG	457	
Qy	362	GGTTTGAGATCGCAGCGCCCAACCATCTTCCCTCCCGCCAGATGGTCAACAAGATGTGC	421	
Db	458	GCCTCTACAGTGGGGCTCTCTCTTGAGCGTGTACGA-----TGGCGCGCTCTTG	509	TGGT
Qy	422	TCATGGCGAAGCCATCATCCACGTGGCGGTCAACTACCGTGTGGCTCTGGGGGTTC	481	
Db	510	ACAGGCCGAGAGACTGTGTGTGTCCATGACTACCGGTGGAGCCTTTGGCTTCCT	569	
Qy	482	GGCTGGTGATGACATCAAGGCCGAGGGCAGCGGGAAACCGCGCTTGAAGCACAGCGTTT	541	
Db	570	GGGCCCTGCCGGGAGC---CGAGAGGCCCGCGGCAATGTGGGTCTCTGTGATCAGAGCCT	626	
Qy	542	GGGCATGCAGTGGTGGCAGACAACATGTGCCGGGTTCCGGCGGGACCCCGAGCAAGGTGAC	601	
Db	627	GGCCCTTCGAGTGGGTGCAGGAGAACTGGCAGCCTTCGGGGGTGACCCCGACATCAGTGAC	686	
Qy	602	ATCTTTGGGAGCGCGGACGA	623	
Db	687	GCTGTTTGGGAGAGCGCGGA	708	

Search completed: August 5, 2004, 17:59:22
Job time : 103 secs

Blank

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 2004, 17:10:06 ; Search time 521 Seconds
(without alignments)
14417.679 Million cell updates/sec

Title: US-09-943-857-3

Perfect score: 1532

Sequence: 1 gccacccgcaagctcgcca.....tgaccaaccgtcttcttg 1532

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 322919 seqs, 2451570024 residues

Total number of hits satisfying chosen parameters: 6445838

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1532	100.0	1532	10	US-09-943-857-3
2	1080.6	70.5	1541	10	US-09-943-857-7
3	1007.6	65.8	1511	10	US-09-943-857-9
4	998.4	65.2	1548	10	US-09-943-857-5
5	830.4	54.2	1469	10	US-09-943-857-1
6	194.8	12.7	1687	16	US-10-369-493-27923
7	88.2	5.8	1716	9	US-09-801-852A-1
8	84.6	5.5	1419	16	US-10-369-493-34501
9	82.8	5.4	1272	16	US-10-369-493-41665
10	61.2	4.0	1476	15	US-10-156-761-7341
11	61.2	4.0	2312	16	US-10-264-049-858
12	61.2	4.0	2428	9	US-09-969-347-220
13	61.2	4.0	2428	10	US-09-418-176-1
14	61.2	4.0	9025608	15	US-10-156-761-1

c	15	60	3.9	740	15	US-10-029-386-22811	Sequence 22811, A
	16	60	3.9	1725	9	US-09-810-861B-5	Sequence 5, Appli
	17	60	3.9	5767	9	US-09-810-861B-3	Sequence 3, Appli
	18	60	3.9	14446	9	US-09-810-861B-4	Sequence 4, Appli
	19	59.6	3.9	3004	10	US-09-873-367C-446	Sequence 446, App
	20	59	3.9	2220	9	US-09-801-852A-3	Sequence 3, Appli
c	21	56	3.7	1126	17	US-10-437-963-58303	Sequence 58303, A
	22	55.8	3.6	1644	13	US-10-425-114-4790	Sequence 4790, Ap
	23	55.8	3.6	3746	15	US-10-160-719-29	Sequence 29, Appl
	24	55.8	3.6	3746	15	US-10-160-719-49	Sequence 49, Appl
	25	55.8	3.6	3773	13	US-10-627-132-9	Sequence 9, Appli
	26	55.8	3.6	3773	15	US-10-209-059-9	Sequence 9, Appli
	27	55.8	3.6	3773	15	US-10-160-719-5	Sequence 5, Appli
	28	53.4	3.5	2191	9	US-09-880-107-3854	Sequence 3854, Ap
	29	53.4	3.5	2191	9	US-09-954-531-1038	Sequence 1038, Ap
	30	53.4	3.5	2191	10	US-09-873-367C-828	Sequence 828, App
	31	53.4	3.5	2191	10	US-09-873-367C-829	Sequence 829, App
	32	53.4	3.5	2484	13	US-09-925-298-271	Sequence 271, App
	33	53.4	3.5	2484	15	US-10-102-806-271	Sequence 271, App
	34	52.2	3.4	1548	16	US-10-369-493-40512	Sequence 40512, A
	35	52	3.4	2092	16	US-10-104-047-249	Sequence 249, App
	36	51.4	3.4	1306	17	US-10-437-963-32084	Sequence 32084, A
	37	51.2	3.3	1728	17	US-10-451-168-40	Sequence 40, Appl
	38	51.2	3.3	1746	13	US-10-114-270-195	Sequence 195, App
	39	51.2	3.3	1746	14	US-10-023-515-3	Sequence 3, Appli
	40	51.2	3.3	1746	17	US-10-674-636-3	Sequence 3, Appli
	41	51.2	3.3	1746	17	US-10-451-168-41	Sequence 41, Appl
	42	51.2	3.3	1857	17	US-10-381-898-20	Sequence 20, Appl
	43	51.2	3.3	1962	16	US-10-094-749-736	Sequence 736, App
	44	51.2	3.3	2158	14	US-10-023-515-1	Sequence 1, Appli
	45	51.2	3.3	2158	17	US-10-674-636-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-943-857-3
; Sequence 3, Application US/09943857
; Publication No. US20030124701A1
; GENERAL INFORMATION:
; APPLICANT: Shaw, Jai-Fu
; APPLICANT: Lee, Guan-Chiun
; APPLICANT: Tang, Shye-Jye
; TITLE OF INVENTION: RECOMBINANT CANDIDA RUGOSA LIPASES
; FILE REFERENCE: 08919-066001
; CURRENT APPLICATION NUMBER: US/09/943,857
; CURRENT FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1532
; TYPE: DNA
; ORGANISM: Candida rugosa
US-09-943-857-3

Query Match 100.0%; Score 1532; DB 10; Length 1532;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GCCACCGCAAGCTCGCCAAACGGCGACACCATCACCGGTCTCAACGGCCATCATCAACGA	60
DB	1	GCCACCGCAAGCTCGCCAAACGGCGACACCATCACCGGTCTCAACGGCCATCATCAACGA	60
QY	61	GGCGTTCCTCGCATTCCTTTGCCAGCGCGGTGGCAACCTTCGGTTCAAGSACCC	120
DB	61	GGCGTTCCTCGCATTCCTTTGCCAGCGCGGTGGCAACCTTCGGTTCAAGSACCC	120
QY	121	TGTGCGTACTCTGGCTGCTCAAGCGGCAGAACTTACTAGCGCGGTGATGAGCAG	180
DB	121	TGTGCGTACTCTGGCTGCTCAAGCGGCAGAACTTACTAGCGCGGTGATGAGCAG	180
QY	181	AACCCCGAGGCGACGTTTGAAGAGAACCTTGGCAAGACGCGACTCGACTTGGTGATGAG	240

```
181 AA CCCCAGGCGAGCTTTGAAGAACTTTGGCAAGCGCACTCGACTGGTGTATGAG 240
241 TCCAAAGTGTTCAGGCGGTGCTTCCCGAGAGTGAAGTGCCTCACCATCAACGTTGGT 300
241 TCCAAAGTGTTCAGGCGGTGCTTCCCGAGAGTGAAGTGCCTCACCATCAACGTTGGT 300
301 CGGCGCGCGGCGCACCAAGGCGGCGCAACCTCCCGGTTCATGCTCTGGATCTTTGGCGGT 360
301 CGGCGCGCGGCGCACCAAGGCGGCGCAACCTCCCGGTTCATGCTCTGGATCTTTGGCGGT 360
361 GGGTTTGAGATCGGCAGCCCCACCATCTTCCCTCCCGCCAGATGGTCAACCAAGTGTG 420
361 GGGTTTGAGATCGGCAGCCCCACCATCTTCCCTCCCGCCAGATGGTCAACCAAGTGTG 420
421 CTATGGGCAAGCCATCATCAGTGGCGGTCAACTACCGGTTCGCTCGTGGGGGTCT 480
421 CTATGGGCAAGCCATCATCAGTGGCGGTCAACTACCGGTTCGCTCGTGGGGGTCT 480
481 TGGCTGTGTGATGATCAAGGCCGAGGCGAGCGGAAACCGCGGTTCGAAGACACAGCGTT 540
481 TGGCTGTGTGATGATCAAGGCCGAGGCGAGCGGAAACCGCGGTTCGAAGACACAGCGTT 540
541 TGGCGATGTCAGTGGGTGGCAGACAATTCGCGGTTCGCGGCGAACCCGAGCAAGTGA 600
541 TGGCGATGTCAGTGGGTGGCAGACAATTCGCGGTTCGCGGCGAACCCGAGCAAGTGA 600
601 CATCTTTGGGAGGCGGCGAGCATGTCGCGGTTCGCGGCGAACCCGAGCAAGTGA 660
601 CATCTTTGGGAGGCGGCGAGCATGTCGCGGTTCGCGGCGAACCCGAGCAAGTGA 660
661 CAACAGTACAGGCAAGCGGTTCGCGGCGGCGATCATGAGGAGCGCATGGTGCC 720
661 CAACAGTACAGGCAAGCGGTTCGCGGCGGCGATCATGAGGAGCGCATGGTGCC 720
721 GGACCCGTTGAGCGGCGAGTACGGCAACGAGATCTACGACCTTTTGTCTGAGTGCTGG 780
721 GGACCCGTTGAGCGGCGAGTACGGCAACGAGATCTACGACCTTTTGTCTGAGTGCTGG 780
781 CTGTGGAGCGGCGAGCGAAGCTCGGTGCTTGGCGAGTGGCGAGCGACCTTGCTCGA 840
781 CTGTGGAGCGGCGAGCGAAGCTCGGTGCTTGGCGAGTGGCGAGCGACCTTGCTCGA 840
841 TGCCACCAACAACTCTCTGGGTCTTGGGTACTCTCTGTTGCGGTGTACTCCGCGCC 900
841 TGCCACCAACAACTCTCTGGGTCTTGGGTACTCTCTGTTGCGGTGTACTCCGCGCC 900
901 CGACGGCAAGAAACATCACCGATGACATGTACAAGTTGGTGGCGAGCGGCAAGTATGCAAG 960
901 CGACGGCAAGAAACATCACCGATGACATGTACAAGTTGGTGGCGAGCGGCAAGTATGCAAG 960
961 GGTTCGCGTGTATCATTTGGGACAGAACGAGGCGGACCATCTTTGGCTCTTGAACGTG 1020
961 GGTTCGCGTGTATCATTTGGGACAGAACGAGGCGGACCATCTTTGGCTCTTGAACGTG 1020
1021 ACCAGAAATGCTCAGGCGGTGCTTACTTCAAGCAGTTTCAACGCCAGCGGAGCGGAG 1080
1021 ACCAGAAATGCTCAGGCGGTGCTTACTTCAAGCAGTTTCAACGCCAGCGGAGCGGAG 1080
1081 ATCGACACTTGAATGGCGGCTGACCCCAAGGACATCAACCGAGGTCGTTGACACGGTT 1140
1081 ATCGACACTTGAATGGCGGCTGACCCCAAGGACATCAACCGAGGTCGTTGACACGGTT 1140
1141 CAAAGCTCACCGGAGTTCAAGAGATCGCGGTGCTCGGCGACCTTGCATTCATCCACG 1200
1141 CAAAGCTCACCGGAGTTCAAGAGATCGCGGTGCTCGGCGACCTTGCATTCATCCACG 1200
1201 CCGCGCGTACTTCTCAACCACTTCCAGGCGGCGACCAAGTACTCGTTCTCAAGCAGC 1260
1201 CCGCGCGTACTTCTCAACCACTTCCAGGCGGCGACCAAGTACTCGTTCTCAAGCAGC 1260
1261 TCGGTTGCCAATCATGGGCACTTCCATGCGCAACGACATGTTGTGGGAGGACTACTTGT 1320
```

```
1261 TCGGTTGCCAATCATGGGCACTTCCATGCCAAGCAATTTGTGTGGAGGACTACTTGT 1320
1321 TGGAAAGCGGCGAGGTCATCTACAACAACCGGTTTATCGGCTTCGCCACCGACTTGGACC 1380
1321 TGGAAAGCGGCGAGGTCATCTACAACAACCGGTTTATCGGCTTCGCCACCGACTTGGACC 1380
1381 CCAACACCGCGGGGTTGTTGTTGAATCGGCCCAAGTACACCCAGCAGCCAGGGCAACAAC 1440
1381 CCAACACCGCGGGGTTGTTGTTGAATCGGCCCAAGTACACCCAGCAGCCAGGGCAACAAC 1440
1441 TGATGATGATCAACCGCTTCGGCTTGATACCCGCAAGGCAACATTCGCAACCGCTGGCT 1500
1441 TGATGATGATCAACCGCTTCGGCTTGATACCCGCAAGGCAACATTCGCAACCGCTGGCT 1500
1501 ACGACGGGTTGATGACCAACCCGTTCTTTGTG 1532
1501 ACGACGGGTTGATGACCAACCCGTTCTTTGTG 1532
```

RESULT 2
US-09-943-857-7
; Sequence 7, Application US/09943857
; Publication No. US20030124701A1
; GENERAL INFORMATION:
; APPLICANT: Shaw, Jai-Pu
; APPLICANT: Lee, Guan-Chiun
; APPLICANT: Tang, Shye-Jye
; TITLE OF INVENTION: RECOMBINANT CANDIDA RUGOSA LIPASES
; FILE REFERENCE: 08919-066001
; CURRENT APPLICATION NUMBER: US/09/943,857
; CURRENT FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1541
; TYPE: DNA
; ORGANISM: Candida rugosa
US-09-943-857-7

Query Match 70.5%; Score 1080.6; DB 10; Length 1541;
Best Local Similarity 84.9%; Pred. No. 7.9e-294;
Matches 1311; Conservative 0; Mismatches 209; Indels 25; Gaps 8;

3 CCACCGCCAAAGCTCGCCAAACCGCGGACACCATACCGGTCTCAACGCCATCATCAACGAGG 62
2 CCCAGCCACGCTCGCCAAACCGCGGACACCATACCGGTCTCAACGCCATCATCAACGAGG 61
63 GTTCTCTGGCATTCCTTTGCGGAGCGCGGTGGGCAACCTCCGCTTCAAGGACCCCTG 122
62 GTTCTCTGGCATTCCTTTGCGGAGCGCGGTGGGCAACCTCCGCTTCAAGGACCCCTG 121
123 TGCCTACTCTGGCTCGCTCAACGGCCAGAAAGTT----ACTTACGGCCCGCTGCATGCAGC 178
122 TGCCTACTCTGGCTCGCTCAACGGCTCAATCTTACCGCGTACGCTCGGTGCATGCAGC 181
179 AGAACCCCGAGGCGACGTTTGAAGAGAACCTTGGCAAGACGGGACATCGACTTGGTATGC 238
182 AGAACCCCGAGGCGACCTACGAGGAGAACCTCCCAAGTGGCGCTTGAATTTGGTATGC 241
239 AGTCCAAAGTGTTCAGGCGGTGCTTCCCGAGAGTGGAGTCCCTCAACCTCAACGTTGG 298
242 AGTCCAAAGTGTTCAGGCGGTGCTTCCCGCAACAGCGAGGACGCTCACCATCAACGTTGG 301
299 TCGCGCCCGCGGCGACCAAGCGGCGGCAACCTCCCGGTCAATGCTCTGGATCTTTGGCG 358
302 TCGCGCCCGCGGCGACCAAGCGGCGGCGCAACCTCCCGGTCAATGCTCTGGATCTTTGGCG 361
359 GTGGGTTTGAAGATCGGAGCGGCCCAACATCTTCCCTCCGCGCCAGATGCTCACCAGAGTG 418
362 GTGGGTTTGAAGATCGGAGCGGCCCAACATCTTCCCTCCGCTCAGATGCTTCCAGAGAGTG 421
419 TGCTCATGGGCAAGCCATCATCAGTGGCGGCGGCAACCTACCGTGTGCTCGTGGGGGTT 478

Db 422 TGCTCATGGGAGCCCATCATCTACGTTGGCGTCAACTACGCTTGGCGTCTTTGGTTT 481
 Qy 479 CTTGGCTGGTGAATGACATCAAGGCGGAGGAGCGGAAAGCGCGCTTGAAGACCAAGCG 538
 Db 482 CTTGGCGCGTCCGACATCAAGGCGGAGGAGCTCAATGCGGCTCAAGGACCAAGCG 541
 Qy 539 TTTGGGATGATGAGTGGTGAGACAAATTTGCGGTTTGGGCGGACCGCAAGGT 598
 Db 542 CTTGGGATGATGAGTGGTGAGACAAATTTGCGGTTTGGGCGGACCGCAAGGT 601
 Qy 599 GACATCTTTGCGAGGCGGAGCATGCTCGTGTGTTGTCACCTCACTGAGCAAGCG 658
 Db 602 GACATCTTTGCGAGGCGGAGCATGCTCGTGTGTTGTCACCTCACTGAGCAAGCG 661
 Qy 659 GACACATGATCAAGGCGGAGCATGCTCGTGTGTTGTCACCTCACTGAGCAAGCG 718
 Db 662 GACACATGATCAAGGCGGAGCATGCTCGTGTGTTGTCACCTCACTGAGCAAGCG 721
 Qy 719 CCGGACCGGTTGAGCGGAGCATGCTCGTGTGTTGTCACCTCACTGAGCAAGCG 778
 Db 722 CCGGACCGGTTGAGCGGAGCATGCTCGTGTGTTGTCACCTCACTGAGCAAGCG 777
 Qy 779 GGTGTGGAGCGGCGGAGCAAGCTCGTGTGTTGTCACCTCACTGAGCAAGCG 838
 Db 778 GGTGTGGAGCGGCGGAGCAAGCTCGTGTGTTGTCACCTCACTGAGCAAGCG 837
 Qy 839 GATGCCAACCAACACTCTCTGGGTTCTTGGCGTACTCTCTGTTGCGTTGTAATCCCG 898
 Db 838 GATGCCAACCAACACTCTCTGGGTTCTTGGCGTACTCTCTGTTGCGTTGTAATCCCG 897
 Qy 899 CCGAGCGGCAAGCAATCAACGATGATGATCAAGTTGTTGCGGAGCGCAAGTATGCA 958
 Db 898 CCGAGCGGCAAGCAATCAACGATGATGATCAAGTTGTTGCGGAGCGCAAGTATGCA 957
 Qy 959 AGGTTCCCGTGATATGCGGAGCAAGCAAGGAGGAGCAATCTTTGGCTCTTGAACG 1018
 Db 958 AGGTTCCCGTGATATGCGGAGCAAGCAAGGAGGAGCAATCTTTGGCTCTTGAACG 1017
 Qy 1019 TGACACGAATGCTAGGCGCGTGTCTTCTTCAAGCAAGTTCTATCCAGCGGAGCGCG 1078
 Db 1018 CCACACCGGAGCGCGGAGGAGCATCTCAGAAAGTTCTATCCAGCGGAGCGCG 1077
 Qy 1079 AGATCAACACTTGTAGTGGCGGATACCCAGGAGATCAACCAAGGTTGCGTTCGACAG 1138
 Db 1078 ATATCAACCACTTGTAGTGGCGGATACCCAGGAGATCAACCAAGGTTGCGTTCGACAG 1137
 Qy 1139 -----TTCAACGCTCACCGCGCTCAAGAGATC---GCGTGTGCGGAGCTTG 1188
 Db 1138 GCATTTCTAAGGCTTTACACCCAGCTCAAGCGGATCAATGCTGTGCTTGGGAGCTCA 1197
 Qy 1189 CATTCATCCAGCGCGCTACTTCTCAACCACTTCCAGGCGGAGCAAGTACTCGT 1248
 Db 1198 CTTTACCTCTCGCGCGCTACTTCTCAACCACTTACCGGTTGCTCCAGTACTCGT 1257
 Qy 1249 TCCTAAGCAGCTCGGTTGCGATCATGGGACCTTCCATGCGCAAGCAATGTGTGGC 1308
 Db 1258 TCCTAAGCAGCTCGGTTGCGATCATGGGACCTTCCATGCGCAAGCAATGTGTGGC 1315
 Qy 1309 AGGACTACTTGTGGAGCGGAGCGTCTATCTACCAACCGGTTTATCGGTTGCGCA 1368
 Db 1316 AGCACTTTTGTGGAGCGGAGCGTCTATCTACCAACCGGTTTATCGGTTGCGCA 1375
 Qy 1369 CCGACTTGAGCCCAACACCGCGGCTTGTGTGAACTGGCCCAAGT--ACACCAAGCAG 1426
 Db 1376 CCGACTTGAGCCCAACACCGCGGCTTGTGTGAACTGGCCCAAGT--ACACCAAGCAG 1432
 Qy 1427 CCAGGCAACACTTGTATGATCAACCGCTTGTGGCTTGTACACCGGCAAGCAACTT 1486
 Db 1433 GCGGGGGGCAACTTGTATGATCAAGTGTGCTTGGCTTGTACACCGGCAAGCAACTT 1492
 Qy 1487 CCGGACCGGCTGATGAGCGGTTGATGACCAACCGGTTCTTTGT 1531
 Db 1493 CCGGACCGGCTGATGAGCGGTTGATGACCAACCGGTTCTTTGT 1537

RESULT 3
 US-09-943-857-9
 ; Sequence 9, Application US/09943857
 ; Publication No. US20030124701A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Shaw, Jai-Fu
 ; APPLICANT: Lee, Guan-Chiun
 ; APPLICANT: Tang, Shye-Jye
 ; TITLE OF INVENTION: RECOMBINANT CANDIDA RUGOSA LIPASES
 ; FILE REFERENCE: 08919-066001
 ; CURRENT APPLICATION NUMBER: US/09/943,857
 ; CURRENT FILING DATE: 2001-08-31
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 9
 ; LENGTH: 1511
 ; TYPE: DNA
 ; ORGANISM: Candida rugosa
 US-09-943-857-9

Query Match 65.8%; Score 1007.6; DB 10; Length 1511;
 Best Local Similarity 85.7%; Pred. No. 2.8e-273;
 Matches 1313; Conservative 0; Mismatches 184; Indels 35; Gaps 16;
 Qy 3 CCACCGCCAGCTCGCCAAAGCGGAGCACCATCAACCGTCTCAAGCGCATCATCAACGAG 62
 Db 2 CCACCGCCAGCTCGCCAAAGCGGAGCACCATCAACCGTCTCAAGCGCATCATCAACGAG 61
 Qy 63 CGTTCCTCGGCATTTCCCTTTGCGAGCGCGGCTGGCAACCTCCGCTTCAAGGACCCCTG 122
 Db 62 CGTTCCTCGGCATTTCCCTTTGCGAGCGCGGCTGGCAACCTCCGCTTCAAGGACCCCG 121
 Qy 123 TGCCTGATCTGCTCGCTCAACCGGCGAGAGTTACTTACGGCCCGTGCATGACAGAGAA 182
 Db 122 TGCCTGATCTGCTCGCTCGATGGCCAGAGTTACTTACGGCCCGTGCATGACAGAGAA 181
 Qy 183 CCGGAGGCGACGTTTGAAGAGAACCTTGGCAGAGCGGCACTCGACTTGGTGTGATGATC 242
 Db 182 CCGGAGGCGACCTTACGAGAGAACCTTCCCAAGGAGCGGCTCGACTTGGTGTGATGATC 241
 Qy 243 CAAGTGTGTTCCAGCGCGTGTCTTCCCGAGAGTGAAGACTGCTTCCATCAACCGTGGTGG 302
 Db 242 CAAGTGTGTTGAGCGGTG-----CCGAGCGAGGACTGTCTCACCATCAACGTTGGTGG 295
 Qy 303 GCGCGCGGACCAAGCGGCGGCGCAACCTCCCGGTGATGCTGTGATCTTGGCGGTTGG 362
 Db 296 GCGCGCGGACCAAGCGGCGGCGCAACCTCCCGGTGATGCTGTGATCTTGGCGGTTGG 355
 Qy 363 GTTTGAGATCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 422
 Db 356 GTTTGAGATCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 415
 Qy 423 CATGGGCAAG--CATCATCCAGTGGCGCTCAACTACCGTGTGCTTCCGTTGGGGGTTCTT 481
 Db 416 CATGGGCAAGCCCATCATCCAGTGGCGCTCAACTACCGTGTGCTTCCGTTGGGGGTTCTT 475
 Qy 482 GCGTGTGATGATCATCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 541
 Db 476 GCGTGTGATGATCATCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 534
 Qy 542 GCGCATGAGTGGGTGGGAGCAAACTTCCCGGCTTCCGCGGCGGCGGCGGCGGCGGCGGCGG 600
 Db 535 GCGCATGAGTGGGTGGGAGCAAACTTCCCGGCTTCCGCGGCGGCGGCGGCGGCGGCGGCGG 594
 Qy 601 CATCTTTGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
 Db 595 CATCTTTGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 654
 Qy 661 CAACACGTACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720
 Db 655 CAACACGTACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 714

```

QY 721 GGACCCGGTGGACGGACGTAACGCAACAGATCTACAGACTCTTTGTCCTCGAGTCTGG 780
Db 715 GGACGC-GTGGACGGCTC-TACGGCAACAGATCTTTGACCTCTGCGCTCGA-CGCGGG 771
QY 781 CTGTGGCAGCGCAGACGACAAAGCTCGCGTGTCTCGCAGTGGCAGCGACACCTTTGCTCGA 840
Db 772 CTGCGGACGGCCAGCAGCAAGCTTGGTGTCTCGCGGTGTGAGCGACAGTGGAGGA 831
QY 841 TGCCACCAACAACACTCTCGGTCTTTGGCGTACTCTCTCGTGGTGGTGTACT-CCCGGC 899
Db 832 CGCACCAACAACACCTCGGTCTTTGGCGTACTCTCTCGTGGTGGTGTACTCTCGCGGC 891
QY 900 CCGACGCAAGAAACATCACCGATGACATGTACAGTTGGTGGCGGACGCAAGTATCAA 959
Db 892 CCGACGCGGTGAACATCACCGACACATGTGC---CTTGGTGGCGAGGCGCAAGTATGCA- 947
QY 960 GCGTTCCTCGTATCATTTGGGACACAGACGAGGCGACCATCTTTGGC-TCTTGAACG 1018
Db 948 ---CTCTGTGTATCATCGGCGACACAGACGAGGCGACCTTCTTTGGCACCTTTGAACG 1004
QY 1019 TGACCAAGATGCTAGCGCCCGTGTACTTCAAGAGATTCATCCAGCGCAGCAGCGGG 1078
Db 1005 TGACCAAGATGCTAGCGCCCGTGTACTTCAAGAGATTCATCCAGCGCAGCAGCGGG 1061
QY 1079 AGATCGACACCTTGATCGCGGCTACCCCGAGGACATCACCGAGGCTCGTTCGACACGG 1138
Db 1062 AG-TGACACAGTTGATGACGCGGTACCCCGAGGACATCACCGAGGCGGTTTCGACACGG 1120
QY 1139 TTCAACG-CTCACCCCGAGTCAAGAGATCGCGTGTCTCGGCGACCTTTGCATTCATC 1196
Db 1121 CTCAAGCCCTCACCCCGAGTCAAGAGATCGCGTGTCTCGGCGACCTTTGCTTACG-- 1178
QY 1197 CAGCCCGCGCTACTCTCAACCACTTCAAGGCGGCGACCAAGTACTCGTTCTCAAG 1256
Db 1179 -----CGTGTCTACTCTCAACCACTTCAAGGCGGCGACCAAGTACTCTCAAG 1233
QY 1257 CAGCTCGGGTGGCAATCATGGGACCTTCCATGCGCAAGCATTTGTGCGAGACTAC 1316
Db 1234 CAGCTCGGGTGGCGGGTCTCGGAACGTTTCACTTCAAGCATTTGTCTTCAGAGCTAC 1293
QY 1317 TTGTGGGAAGCGGCGAGCGTCTATCTACAAACAGCGTTTATCGCGTTCGCCACCGATTG 1376
Db 1294 TTGTGGGACGGCTCGCTCATCTACAAACAGCGTTTATCGCGTTCGCCACCGATTG 1353
QY 1377 GACCCCAACACCGCGGGTGTGTGTGAACCTGGCCCAAGTACACAGCAGCAGCGGCAAC 1436
Db 1354 GACCCCAACACCGCGGGTGTGTGTGAAGTGGCCCGAGTACACAGCAGCAGCGGCAAC 1413
QY 1437 AACTTGATGATCATCAACGCTTGGCTGTGTACCGCGCAAGGACAACTTTCGACCGCT 1496
Db 1414 AACTTGATGATCATCAACGCTTGGCTGTGTACCGCGCGAGTGTACCGCGGAGACAACT- 1472
QY 1497 GGCTACGAGCGGTTGATGACCAACCGGTTCTT 1528
Db 1473 GGCTACGAGCGGTTGTTCTTCAACCGCGGTT 1504

```

RESULT 4

```

US-09-943-857-5
; Sequence 5, Application US/09943857
; Publication No. US20030124701A1
; GENERAL INFORMATION:
; APPLICANT: Shaw, Jui-Fu
; APPLICANT: Lee, Guan-Chiun
; APPLICANT: Tang, Shye-Jye
; TITLE OF INVENTION: RECOMBINANT CANDIDA RUGOSA LIPASES
; FILE REFERENCE: 08919-066001
; CURRENT APPLICATION NUMBER: US/09/943,857
; CURRENT FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5

```

```

; LENGTH: 1548
; TYPE: DNA
; ORGANISM: Candida rugosa
US-09-943-857-5

Query Match      65.2%; Score 998.4; DB 10; Length 1548;
Best Local Similarity 83.8%; Pred. No. 1.1e-270;
Matches 1301; Conservative 0; Mismatches 221; Indels 30; Gaps 14;

QY 1 GCCACCGCAAGCTCGCCAAACGCGACACCATCACCGGTCTCAACGCCATCATCAACGA 60
Db 1 GCCACATGCCACGCTCGCCAAACGCGACACCATCACCGGTCTCAACGCCATCATCAACGA 60
QY 61 GGGGTTCTCTCGGATTCCTTTTGGCGAGCCGCGGTGGGCAACCTCTCGCTTCAAGAGCCC 120
Db 61 GGGGTTCTCTCGGATTCCTTTTGGCTAGCCGCGGTGGGCAACCTCTCGCTTCAAGAGCCC 120
QY 121 TGTGCGGTACTCTGCGTCTCAACGCGCAGAGTT---ACTTACGCGCGGTGATGCGAG 177
Db 121 TGTGCGGTACTCTGCGGCTCTCAATGCTCAGAGTTTACTGTATGGCGCTGTGATGCGAG 180
QY 178 CAGAACCCCGAGGCGACGTTTGAAGAAACCTTGGCAAGACGCGACCTCGACTTGGTGATG 237
Db 181 ATGAACCAATTGGCAACTGGGACTCTCTCGTTCCCHAGGCTGCATCAACT---TGATG 237
QY 238 CAGTCCAAAGTGTTCAGGGCGGTCTTCCCGAGAGTGAAGACTGCTTCAACCATCAAGTG 297
Db 238 CAGTCCAAAGTCTTTCAGGGCGGTCTTCTTAAACGCGGAGGACTGTCTCAACCATCAAGTG 297
QY 298 GTGCGCGCGCGGACCAAGCGGCGGCGCAACCTCCCGGTCTCTGATCTTTTGGC 357
Db 298 GTGCG---GCCGCGCACCAAGCGGCTGCCAACCTCCCGGTGATGTTGTTTGGC 354
QY 358 GGTGGTTTGAAGTTCGCGAGCGCCCAACCTCTTCCCTCCCGCCAGATGGTCAACAGAGT 417
Db 355 GCGCGGTTTGAAGTTGGCGGCTCCAGTCTCTTCCCTCCCGCACAGATGATCAGCGCCAGC 414
QY 418 GTGCTCATGGCAAG-CGATCATCAAGTGGCGGTCAACTACCGTGTTCCTCTGTTGGGG 476
Db 415 GTGCTTATGGCAAGCGCCATCATCAAGTGAAGTGAATACCGGTTGTTCTGTTGGGG 474
QY 477 TCTTGGTGTGATGATCATCAAGCGGCGGCGGAGGAGCGCGGTCTGAAGACCGAG 536
Db 475 TCTTGGTGTGTCAGATCATCAAGCGCGGCGGAGGAGGAGCGCGGTGTCAGACCGAA 534
QY 537 GGTGGGCGATGCGGTGGCGACAAACATTTGCGCGGTTTCGGCGGCGACCGGAGCAAG 596
Db 535 CGTTGGGTTTTCAGTGGGTGGCGCAACATTTGCGCGGTTTCGGCGGCGACCGGTTCCAG 594
QY 597 GTGA-CATCTTTTGGCGA-GGCGGGCAGCATGTCGCTGTGTGCGCACTCATCTGGAACGA 654
Db 595 GTGACCATCTTTTGGTGGCGGCGGCGAGCATGTCGCTGAATGTGTCAAGTCTCTCTGGAACGA 654
QY 655 CGGCGCAACAGCTTACAGGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 714
Db 655 CGGCGCAACAGCTTACAGGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 714
QY 715 GGTGCG-GGACCGCGTGGACGCGCATGTCAGGCAACAGAGATCTACGACCTCTTTGTCTGA 773
Db 715 GGTGCGGCGGACCGCGTGGATGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 774
QY 774 GTGCTGGTGTGGCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 832
Db 775 CAGCGCGGTGTGGCAGTGCAGCGCAAGCTCGGCTGTGTGGCGAGCATGTCGAACGACAA 834
QY 833 TTGCTCGATGCCACCAACAAACACTCTCTGGTCTTTGGCGTACTCTCTGTTGGCGGTTG 889
Db 835 CTCCTCCAGGCGCACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 894
QY 890 -TACTCCCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 948
Db 895 TCTCCCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 954

```

```
QY 949 CAAGTATGCAAGCTTCCCGTGATCATTTGGCGAACAGAGGCAACCATCTTTG- 1007
Db 955 CAAGTGTGCCAACCTTTCCGGTGATCATTTGGCGAACAGAGGCAACAGTGTTCG 1014
QY 1008 GCTCTTGAAGTGAACGATGCTCAGGCGGTGCTTACTTCAAGCAGTTTCAATCCAGC 1067
Db 1015 GTTGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1074
QY 1068 CAGCGACGCGGAGATGACACCTTTGATGGCGGCGTACCCCGAGGACATCAACCCAGGGTCC 1127
Db 1075 CAGCGACGCGGAGATGACACCTTTGATGGCGGCGTACCCCGAGGACATCAACCCAGGGTCC 1134
QY 1128 GTTTCGACAGG-----TTCAACGC--TCACCCCGAGTTCGAAGAAT---CGCGGTGCT 1177
Db 1135 GTTTCGACAGGCGCATCTTCAACGCATCAACCCCGAGTTCGAAGAAT---CGCGGTGCT 1194
QY 1178 CGCGACCTTTGATTCATTCACGCGCGCGGTACTTCTCAACCACTTCCAGGCGGCGAC 1237
Db 1195 TGGTGACCTTGGTTCCTCTCCCCCGCGCTACTTCTCAACCACTTCCAGGCGGCGAC 1254
QY 1238 CAAGTACTCGTTCTC--AAGCAGCTCGGGTTGCCAATCATGGGCACTTCCATGCCAACG 1296
Db 1255 CAAGTACTCGTTCTCGAAGCAGCTTGGTTGCCGCTGATGGACCCCAACG 1314
QY 1297 ACATTTGTGGAGGACTACTTGTGGGAAGCGGCGATCATCTACAACACGCGTTTA 1356
Db 1315 ACATTTGTGGAGGACTTTTGTGGAGCGGCGATCATCTACAACACGCGTTTA 1374
QY 1357 TCGGTTTCGACCGACTTCGACCCCAACCGCGGGGTTGTTGGTGAACCTGGCGCAAGT 1416
Db 1375 TTGCCTTGGCAAGCCTCGACCCGGAACAGGCGGTTTGTGTGAACCTGGCGCAAGT 1434
QY 1417 ACACGACGACGAGGCAACAACTTGATGATGATCAACGCTTGGGCTTGTACACCGGCA 1476
Db 1435 ACACGACGACGAGGCAACAACTTGTTCAGATCAACGCTTGGGCTTGTACACCGGCA 1494
QY 1477 AGACAACTTCGCGACCGCTGGGTACGACGCTGTGATGACCAACCCGTTCTT 1528
Db 1495 AGACAACTTCGCGACCGCTGGGTACGACGCTGTGATGACCAACCCGTTCTT 1546

RESULT 5
US-09-943-857-1
; Sequence 1, Application US/09943857
; Publication No. US20030124701A1
; GENERAL INFORMATION:
; APPLICANT: Shaw, Jai-Fu
; APPLICANT: Lee, Guan-Chiun
; APPLICANT: Tang, Shye-Jye
; TITLE OF INVENTION: RECOMBINANT CANDIDA RUGOSA LIPASES
; FILE REFERENCE: 08919-066001
; CURRENT APPLICATION NUMBER: US/09/943,857
; CURRENT FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1469
; TYPE: DNA
; ORGANISM: Candida rugosa
US-09-943-857-1
```

```
Query Match 54.2%; Score 830.4; DB 10; Length 1469;
Best Local Similarity 78.1%; Pred. No. 2.1e-223;
Matches 1202; Conservative 0; Mismatches 241; Indels 97; Gaps 13;

QY 10 CAAGTTCGCCAACGGGACACCATACCGGTCTCAACGGCATCATCAACGAGCGTTCCT 69
Db 8 CCACCTCGCCAACGGGACACCATACCGGTCTCAACGGCATGTCAACGAAAAGTTTCT 67
QY 70 CGGCATTCCCTTTGCCAGCGCGGTGGGACACTCCGCTTCAAGGACCTGTGCGGTA 129
Db 68 CGGCATTCCCTTTGCCAGCGCGGTGGGACA--CTCCGCTTCAA-----110
```

```
QY 130 CTCTGGCTCGCTCAACGGCCAGAA--TTACTTACGGCCCGTGATGACAGCAAGACCCGA 188
Db 111 -----GCTCAACGGCCAGAGTTTACTTACGGCCCGTGATGACAGATGAACCTTAT 161
QY 189 GGGCACTTTGAAGAGAACCTTTGGCAAGACGGCACTCGACTTTGGTGTGATGCAAGGT 248
Db 162 GGGCTCGTTTCA-----TTGGTGTCTCCAGTCCAAGAT 193
QY 249 GTTCCAGCGGTGTTCCCGAGAGTGAAGTGCCTTCAACATCAACGTGTGCGCGCGCC 308
Db 194 CTTTCAAGTGTGTTTCCCAACGACGAGGACTGTCTCAC-----232
QY 309 GGGCACCAAGCGCGGCCCAACCTCCCGGTCTTCTGGATCTTTGGCGGTGGTTGA 368
Db 233 ---CACGAGGCGAGTGTGCTCTCCCGGTGATGCTCTGGATCTTTGGCGGTGGTTGA 289
QY 369 GATCGGAGCGCCCAACCATCTTCCCTCCCGCCAGAGTGTCAACAGAGTGTCTCATGGG 428
Db 290 GCTTGGCGGTCTCAGCCTCTTCCAGGAGACAGATGTTGGCCAAAGAGCGTGTCTATGGG 349
QY 429 CAGGCC--ATCATCCAGTGGCGGTCAACTACCGTGTTCCTCTGTTGGGGGTCTTTGGCTGG 487
Db 350 TAAACCGGTGATCCACGTGAGCATGAATACCGGTGGCGTCTATGGGGGTCTTTGGCGGG 409
QY 488 TGATGACATCAAGCGCGGAGCGGCAACGAGGCGGAGCGGCTTGCATGACGAGCGCTTGGCCAT 547
Db 410 CCCCAGCATCCAGAACGAGGCGGAGCGGAGCGGCTTGCATGACGAGCGCTTGGCCAT 469
QY 548 GCGTGGGTGGGAGAGCAAACTTCCCGGTTCCGCGGCGACCCGAGCAAGGTGA--CATCTT 606
Db 470 GCAGTGGGTGGGAGCAAACTTCTGGGTTTGGCGGCGACCCGAGCAAGGTGACCATATA 529
QY 607 TGGCGAGGCGGCGAGCATGTCCGTTGTGCGACCTCATCTGAAACGACGCGGCAACAC 666
Db 530 CGCGAGGCGGCGAGCATGTGCGACCTTGTGTGCACTTGTGTGAAACGACGCGCAACAC 589
QY 667 GTAAAGGGCAAGCGGTTTCCGCGGCGCATCATGACGAGGAGCGCATGTGCGGACCC 726
Db 590 GTAAAGGGCAAGCGGTTTCTCCGCGCGCATCATGACGAGGAGCGCATGTGCGGACCC 649
QY 727 GTTGAAGCGGACGTACGGCAACGAGATCTACGACCTCTTTTGTCTCGAGTCTGCTGTGG 786
Db 650 GGTGAGCGGACGCTACGGCACCGAGATCTACAAACGAGTGTGCGGTCTGCGGGTGG 709
QY 787 CAGCGCGGAGCAAGCTGCGTGTGCGCAG---TGAGGAGCACCTTGTCTGATGC 843
Db 710 CAGTGCAGCGACAACTTCCGCTTGGCGGCGCTTCTCAGGACACGTTGTACCGAGC 769
QY 844 CACCAACAACTCTCTCGGTTCTTGGCGTACTCTCTGTTGCGGTTGTA---CTCCCGGCC 900
Db 770 CAGGAGCGGACACCGCGCGGTGTTGGCGTACCGCTGTTGCGGTTGTAICTCCCGCGGCC 829
QY 901 CGACGCAAGAAACATCACCGATGACATGTAAGTTTGGTGGCGGCGGCAAGTATCAAG 960
Db 830 CGACGCGACCTTCAACCGGACGACATGTATGCTTGGTGGCGGCGGCAAGTATCAAG 889
QY 961 CGTTCCTGATCATTGGGCGGACGAGACGACGAGGCGACATCTTT--GGCTTTGAACGT 1019
Db 890 CGTTCCTGATCATTGGGCGGACGAGAACGAGGCGGACCTTTTGTGGGCTCTTGAACGT 949
QY 1020 GACCGAATGTCTCAGGCGCGGTCTTACTTCAAGCAGTTTCAACGCGGCGGCGGCA 1079
Db 950 GACCGAATGTCTCAGGCGGCGGTCTTCAAGCAGTTTCAACGCGGCGGCGGCA 1009
QY 1080 GATCGACACTTGTATGGGCGGTACCCCGAGGACATCAACCGGCTTCTCGACACG-- 1138
Db 1010 GATCGACACTTGTATGGGCGGTACACGAGGACATCAACCGGCTTCTCGACACG-- 1069
QY 1139 -----TTCAACGCTCAACCGGAGTTCAAGAGATGCGGCTGTGCGGCACTTGCATT 1192
Db 1070 CATCTTCAATGCATCAACCGGAGTTCAACCGGAGTTGCGGCTGTGCGGCACTTGCATT 1129
QY 1193 CATCCAGCGCGCGCTACTTCTTCAACCACTTCCAGGCGGCGGCAACGAGTCTGCTCT 1252
```


Db 1130 CAGCTTGGCGTCCGCTCTCTCTCACTACTACAGGGCGCACCAAGTACTCGTT-CT 1188
Qy 1253 CAAGCAGCTCGGTTGCAATCATATGCGCACTTCCATCGCCAAAGCAATGTGTGGAGGA 1312
Db 1189 CAAGCAGCTTGGTTGGCGCTTGGCGCACTTCCAGCGCAACGACATCATCTGGCAGGA 1248
Qy 1313 CTACTTGTGGGAAGCGGACGCTCATCTACAACAGCGTTTATCGGTTTCGCCACCGA 1372
Db 1249 CTACTTGTGGGACGCGGAGTGTGATCTACAACAGCGTTTATCGGTTTGCACCGA 1308
Qy 1373 CTTGCAACCCCAACACCGCGGTTTGTGGTCAACTGGCCCAAGTACACAGCAGCCAGGG 1432
Db 1309 CCTCGACCCGCAACAGCGGCTTGTGACCACTAGCGCCACGATACACAGCAG-CAGGG 1367
Qy 1433 CAACAACCTTGATGATGATCAACGCTTGGGCTTGTATACCGGGCAAGGACAACTTCCGCAC 1492
Db 1368 CAACAACCTTGATGATGATCAACGCTTGGGCTTGTATACCGGGCAAGGACAACTTCCGCC 1427
Qy 1493 CGTGGCTACGACGCTTGTATGACCAACCGCTTCTTTGTG 1532
Db 1428 GGATGCTGATACGCGCTCTTTTCCAACCGCTTCTTTG 1467

RESULT 6

US-10-369-493-27923
; Sequence 27923, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 27923
; LENGTH: 1687
; TYPE: DNA
; ORGANISM: Neurospora crassa
US-10-369-493-27923

Query Match 12.7%; Score 194.8; DB 16; Length 1687;
Best Local Similarity 55.6%; Pred. No. 2e-44;
Matches 589; Conservative 0; Mismatches 392; Indels 79; Gaps 8;
Qy 25 CGACACCATCACCGGTTCTCAAGCCATCATACAGAGGGCTTCTCGGCAATCCCTTTCG 84
Db 27 CGGTACGATTGTGGTGTAAACGGCATCTCTACTAGAGGCTTCAACGGAATCCCTACGC 86
Qy 85 CGAGCCCGCGGTGGCAACTCCGCTTCAAGGACCTGTGCGGTACTCTGGCTCGCTCAA 144
Db 87 CTTCTCCGACCGCAACCTTCGCTCAAGCTCCCGTGAGACTTAAGTGTCTCTGG 146
Qy 145 CGGCCAGAAAGTTACTTA-----CGGCCGCTGCATGCAGCAGAAACCCGAGGGCAGTTGA 200
Db 147 TGTCTTTGATGCGTCTGGCATCGGCTTGTGTCGCCCGGCTTCTTGTGACACCTCGTC 206
Qy 201 AGAAGAACCTTGGCAAGACGCACTCGACTTGTGTGATGAGTCAAGGTGTTCAGGCGGT 260
Db 207 GAACGAGTTTCTGCTCAGGTATCGATAAGATCGTTAAACACGACGCTTTTCAAGACTAT 266
Qy 261 GCTTCCCGCAGTGAAGTACGCTTCCATCATCAAGTGTGCGCGCGCGGCAAGGC 320
Db 267 ACTCAACGCTCAAGAGACTGCTTGAACCATCTCGGTCACTCGTCCCAAGGCAAGGC 326
Qy 321 GGGCGCCAACTCTCCCGCTCATGCTCTGGATCTTTGGCGGTGGTTG----- 367

RESULT 7

US-09-801-852A-1
; Sequence 1, Application US/09801852A
; Patent No. US20020048781A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; APPLICANT: Chiang, Shu-Jen
; APPLICANT: Jonathan, Basch
; TITLE OF INVENTION: DIRECT PRODUCTION OF DESACETYLCEPHALOSPORIN C
; FILE REFERENCE: ON0163
; CURRENT APPLICATION NUMBER: US/09/801,852A
; CURRENT FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/188,033
; PRIOR FILING DATE: 2000-03-09
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1716
; TYPE: DNA
; ORGANISM: Rhodosporidium toruloides
US-09-801-852A-1

Db 327 TGGTGATAAGCTCCCGTCTCTTTTCTGGATCTTTTGGTGGTGGTTTCAAGTGAGAAATCC 386
Qy 368 -----AGATCGGACAGCCCAACA 385
Db 387 AGCTTATATAGCGGATGAATGAACAGTGTCTCAACAGCTCGGATCGGCGTCCA 446
Qy 386 TCTTCCCTCCCGCCAGATGTCACCAAGAGTGTCTCATGGGCAAGCCAT-CATCCACG 444
Db 447 TGTACGATGGCGCTCCCTAGTCACCAACGCTATCAACAGGTAAGCCGTACGCTTACG 506
Qy 445 TGGCCGTCAACTACGCTGTCCTCGTGGGGTCTTGGCTGTGTGATGATCAATCAAGGCG 504
Db 507 TTGCGGTCAACTACGCTGTCGCTGGCTTTTGTGTTTCATGCCGGAAGAGATCTTAAG 566
Qy 505 AGGCGAGCGGAAACCGCGCTTGAAGGACCAAGCGTTTGGGCATGCAAGTGGTGGCAGCA 564
Db 567 ACGGCTCTTCCAACTTGGGTCACTTGAACGAGCGCATGGGCTCCAGTGGGTGGCGACA 626
Qy 565 ACATTGGCGGTTCCGGCGGACCCGAGCAAGGTGAC-ATCTTTGGCGAG--GCGGCA 620
Db 627 ACATTGCTGCTTCCGCGGTGACCCAGCAAGGTCACTATCTGGGGCGAGTCCGCGGTG 686
Qy 621 GCATGTCGCTGTTGTGCCACCTCATCTGGAACGAGCGGCAACACAGTACAGGGCAAGC 680
Db 687 CCATGTCGCTTCAACCGAGTGTCTCTATGACGGTGAACAACAGTACAAAGCAAGC 746
Qy 681 CGTTGTTCCGCGCGGATCATGAGGGAG-----CCATGTCGCGGACCCCGTGGAGC 734
Db 747 CCCTTTCCGTGGCGCATCATGAATCTGCTTCCATCGTCCCGCGCGCTCGACT 806
Qy 735 GCAGTACGGCAACAGATCTTACGACCTTCTTGTCTCAGTGTCTGGCTGTGGCAGGCA 794
Db 807 GCCCAAGGGCCAGAAAGTCTACGACACCGTCTCAAGAAGCCCGGCTGCTCTGGTGTG 866
Qy 795 GCGCAAGCTCGCGTGTGGC-----AGTGGAGCGACACCTTGTCTCATGCCACCAACA 851
Db 867 CTGACACCTTGTCTTGGCTTGGCGCTTCTCCCTAGACACTTCTTCAAGGCGCTAAT 926
Qy 852 ACACCTCTGCGGTTCTTGGCTACTCTCTGTTGCGGTTGT-----ACTCCCGGCGGAGC 905
Db 927 CGTGTCTGGGATCTCTGCTACAACTCCGTGCTCTTTTACCTCCCGACCGCATG 986
Qy 906 GCAAGAACATCACCGATGACATGTACAGTTGGTGGCGACCGCAAGTATCAAGCGTTC 965
Db 987 GCAAGGCTTGACTCAGAGCGCGGATAAGCTCATGCTCGCTAAGAGTACGCGCGCTCC 1046
Qy 966 CGTGATCATTTGCGGACAGCAAGCGAGGCGACCATCTT 1005
Db 1047 CCATGATCATCGGCGATCAAGAGATGAGGCGACTCTCTT 1086


```
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7341
; LENGTH: 1476
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1476)
US-10-156-761-7341

Query Match 4.0%; Score 61.2; DB 15; Length 1476;
Best Local Similarity 53.0%; Pred. No. 7.7e-07;
Matches 215; Conservative 0; Mismatches 168; Indels 23; Gaps 3;

QY 216 GACGCACTCGACTGGTGTGATCGATCCAAAGGTGTTCCAGCGGTGTTCCCGAGTGA 275
DB 144 GCGGCCCCAGGAGACGGGGATCCAGGGCGCGGCCCTGTGTGACGGCGCCACGGGCGA 203
QY 276 GGACTGCTCAACATCACTGCTGCGCGCGCGGACCAAGGGGGGCGCACTCC 335
DB 204 CGACTGGTGTGAGTCTACGGCGGTGCTTACAGCTCGGCCACTCGGCGCGCGGCTA 260
QY 336 GGTATGCTCTGGATCTTTGGGGTGGTGTGAGATCGGACCGCCACCATCTTCCCTCC 395
DB 261 GGTATGTTGTGATCTACGGCGGTGCTTACAGCTCGGCCACTCGGCGCGCGGCTA 320
QY 396 CCGCCAGATGTTCAACAGAGTGTGTCATGCGGCAAGCCATCATCAAGTGGCGGTCAAC 455
DB 321 CGACGCTCGCGCATC-----GCCGCGACGGCGACGTCGTGTGTCAACCTCAAC 372
QY 456 TACCGTGTGCTGTTGGGGTGTGTTGGGCATGTCAGTGGGTGGCGACACATTCGCGGG 515
DB 373 TACCGTGTGCGCATCGAGGGCTTCGCGCGGTGTCGACGGCGCTCCCGCC----- 420
QY 516 AACGCGGCTTCAAGCACGAGCTTTGGGCATGTCAGTGGGTGGCGACACATTCGCGGG 575
DB 421 AACCGCGTCTGTCGACCAAGTTCGCGCGGTGATGATGATGATGATGATGATGATGATG 480
QY 576 TTCCGCGCGCACCCCGAGCAAGTGTGATCTTTGGAGCGCGGGCAG 621
DB 481 TTCCGCGCGCACCCCGCGCGGTACCGTCTTCGGGGAGTCGCGCG 526

RESULT 11
US-10-264-049-858
; Sequence 858, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P133P1
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
```

```
; SEQ ID NO 858
; LENGTH: 2312
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1763)..(1763)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-264-049-858

Query Match 4.0%; Score 61.2; DB 16; Length 2312;
Best Local Similarity 52.7%; Pred. No. 8.7e-07;
Matches 214; Conservative 0; Mismatches 173; Indels 19; Gaps 3;

QY 236 TGCAGTCCAAAGGTGTTCCAGGCGGTGTTCCCGAGTGAAGTCTGCTCACCATCAACG 295
DB 267 TGCAGGCCACCATCACCAGGACAGCACCTACGGGGATGAAGACTGCTGTACCTCAACA 326
QY 296 TGGTGGCGCGCGCGGACCAAGCGGGGGCC---AACCTCCCGGTGATGCTCTGGATCT 352
DB 327 TTTGGGTGCGCGCGAGGAGCAAGTCTCCCGGACCTCCCGGACCTGCTGATGATCTGGATCT 386
QY 353 TTGGCGGTGGTGTGAGATCGGCAGC-----CCCAACATCTTCCCTCCCGCCCGAG 402
DB 387 ATGGAGGCGCTTCTCTCATGGGTCCGGCCATGGGGCAACTCTCTCAACAACATACCTGT 446
QY 403 ATGCTCACCAGAGTGTCTCATGGGCAAGCCATCATCCAGTGGCGGTCACTACCGTG 462
DB 447 ATGACGGCGAGGATGCCACACGCGAAGCGTCTGCTGCTCACCCTTCACTACCGTG 506
QY 463 TTGGCTGTGGGGTCTTGGCTGTGATGATCAAGCGGAGGCGGAGCGGAAACGCG 522
DB 507 TCGGCGCGCTTGGGTTCCTCAGCACTGGGGAC-----GCCAATCTGCAGGTACTATG 560
QY 523 GCTTGAAGACACGAGTGTGGGCATGCACTGGGTGGCAGCAACAATTCGCGGTTCGCGG 582
DB 561 GCCTTGGGATCAGCAGATGCCCATGCTTGGGTGAAGAGGAATATCGCGGCTTCGCGG 620
QY 583 GCGACCCGAGCAAGTGTGATCTTTGGGAGCGGCGGAGCATGTCC 628
DB 621 GGGACCCCAACAACATCACGCTTTCGGGGAGTCTGCTGGAGGTCC 666

RESULT 12
US-09-969-347-220
; Sequence 220, Application US/09969347
; Patent No. US20020115085A1
; GENERAL INFORMATION:
; APPLICANT: Ebner, Reinhard
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 689290-69
; CURRENT APPLICATION NUMBER: US/09/969,347
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US/60/237,598
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: US/60/237,604
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 318
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 220
; LENGTH: 2428
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-969-347-220

Query Match 4.0%; Score 61.2; DB 9; Length 2428;
Best Local Similarity 52.7%; Pred. No. 8.8e-07;
Matches 214; Conservative 0; Mismatches 173; Indels 19; Gaps 3;

QY 236 TGCAGTCCAAAGGTGTTCCAGGCGGTGCTTCCCGAGTGAAGTCTGCTCACCATCAACG 295
DB 344 TGCAGGCCACCATCACCAGGACAGCACCTACGGGGATGAAGACTGCTGTACCTCAACA 403
```



```

; LOCATION: 2152..2184
; FEATURE:
; NAME/KEY: repeat unit
; LOCATION: 2185..2217
; FEATURE:
; NAME/KEY: repeat unit
; LOCATION: 2218..2250
; FEATURE:
; NAME/KEY: repeat unit
; LOCATION: 2251..2283
; PUBLICATION INFORMATION:
; AUTHORS: Nilsson, Jeanette
; AUTHORS: Blackberg, Lars
; AUTHORS: Carlsson, Peter
; AUTHORS: Enerback, Sven
; AUTHORS: Hernell, Olle
; AUTHORS: Bjursell, Gunnar
; TITLE: cDNA cloning of human-milk
; TITLE: bile-salt-stimulated lipase and evidence for its
; TITLE: identity to pancreatic carboxylic ester hydrolase
; JOURNAL: Eur. J. Biochem.
; VOLUME: 192
; PAGES: 543-550
; DATE: Sept.-1990
;
US-09-418-176-1

```

```

Query Match 4.0%; Score 61.2; DB 10; Length 2428;
Best Local Similarity 52.7%; Pred. No. 8.8e-07;
Matches 214; Conservative 0; Mismatches 173; Indels 19; Gaps 3;

Qy 236 TGCAGTCCAAAGGTGTTCCAGGCGGTGTTCCCCAGAGTGAGGACTGCTCCACCATCAACG 295
Db 344 TCGAGGCCACCATCACCAGGACAGCACCTAGCGGGATGAAGACTGCTGTACCTCAACA 403
Qy 296 TGGTGGCGCGCGCGGACCAAGAGCGGCGCC---AACTCCGGTCAATGCTCTGGATCT 352
Db 404 TTTGGGTGCCCCAGGCGGAGCAAGTCTCCCGGACCTCCCGGCTTATGATCTGGATCT 463
Qy 353 TTGGCGGTGGGTGATGATCGGCAGC-----CCACATCTTCCCTCCCGCCAG 402
Db 464 ATGGAGGGCGCTTCTCTATGGGTCCGGCATGGGGCAACTTCTCAACAACTACCTGT 523
Qy 403 ATGCTCACAAGAGTGTCTCATGGGCAAGCCATCATCCAGTGGCGGTCAACTACCGTG 462
Db 524 ATGACGGCGAGAGATGCCACACGCGGAAGCTCATGCTGTGATCACTTCAACTACCGTG 583
Qy 463 TTGCTCTGGGGTCTTGGTGTGATGATCAAGCCGAGGCGAGCGGACGCGG 522
Db 584 TCGGCCCGCTTGGGTCTCTCAGCACTGGGAC-----GCCAATCTGCCAGGTAATATG 637
Qy 523 GCTTGAAGGACGAGGTTTGGGATGCGATGCGGTGGGCGAGACAACTTCCGGGTTCGGCG 582
Db 638 GCCTTCGGGATCAGACATGGCCATGCTTGGGTGAAGAGNATATCGGGCTTCGGGG 697
Qy 583 GCGACCCGAGGAGGTGATCATCTTTGGCGAGCGCGGCGAGCATGTCC 628
Db 698 GGGACCCCAACAACATCAGCTCTTCGGGGAGTCTGCTGGAGGTGC 743

```

```

RESULT 14
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262

```

```

; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 4.0%; Score 61.2; DB 15; Length 9025608;
Best Local Similarity 53.0%; Pred. No. 7.4e-06;
Matches 215; Conservative 0; Mismatches 168; Indels 23; Gaps 3;

Qy 216 GACGGCACTCGACTTGGTGTGATGCAGTCCAAAGGTGTTCCAGGCGGTGCTTCCCAAGTGA 275
Db 8791846 GCCGCCCCAGGAGACGGGATCCAGGGCGCGCGCTGCTGGACGCGCCACACGGCGA 8791787
Qy 276 GGACTGCTCACCATCAACGTGGTGGCGGCCCGCGGCACCAAGGGGGCGGCCAACCTCCC 335
Db 8791786 CGACTGGTGTACGGTCAAC---GTCTGGACACACCGGACCGCTGTGCGCCGCGGCC 8791730
Qy 336 GGTATGCTCTGGATCTTTGGCGGTGGTGTGATCGGACGCCACCATCTTCCCTCC 395
Db 8791729 GGTATGCTGTGATCTACGGCGGTGCTACAAGCTCGGCACATCCGCGACCCCGGTA 8791670
Qy 396 CGCCAGATGTACCAAGAGTGTCTCATGGGCAAGCATATCCAGTGGCGGTCAAC 455
Db 8791669 CGAGCTCGCGCATC-----GCCGCGACGCGGACGTGCTGTCACTCAAC 8791618
Qy 456 TACGCTGTGCTCGTGGGGTCTTGGCTGTGATGATCAAGCGGCGAGGGCGAGCGGG 515
Db 8791617 TACGCTGTGCGCATCGAGGGCTTCGCCCGGTTCGACGCGCTCCCGCC----- 8791570
Qy 516 AACCGCGCTTGAAGGACCGGTTTGGCATGCGATGGGTGGGAGACAACTTCCCGGG 575
Db 8791569 AACCGCGTCTGCTCGACCGGCTCGGCGCTTGAATGGGTACGGGAGAACATCAGCG 8791510
Qy 576 TTGCGCGCGACCGCGAGCAAGGTGACATCTTTGGCGAGCGGGCGAG 621
Db 8791509 TTGCGCGCGACCGCGCGGCTCACCGTCTTGGGGAGTCGGCG 8791464

```

```

RESULT 15
US-10-386-22811/c
; Sequence 22811, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEONICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 22811
; LENGTH: 740
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF002993.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7

```

Job time : 538 secs

Blank

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 2004, 15:42:35 ; Search time 2825 Seconds
(without alignments)
16194.287 Million cell updates/sec

Title: US-09-943-857-3

Perfect score: 1532

Sequence: 1 gccaccgcgaagctcgcca.....tgaccacccgttttctgtg 1532

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_pbg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	173.2	11.3	769	14	CF881106
C 2	173.2	11.3	824	14	CB907899
C 3	159.2	10.4	719	14	CF880958
C 4	159.2	10.4	774	14	CB907713
					CF881106 tric083xn
					CB907899 tric083xn
					CF880958 tric083xc
					CB907713 tric083xc

C	5	150.6	9.8	970	14	CF885800
	6	117.4	7.7	1070	14	CF885724
	7	110	7.2	708	14	CF880590
	8	110	7.2	762	14	CB907475
	9	104.8	6.8	763	29	CG810164
	10	88.6	5.8	681	14	CD054203
	11	86.4	5.6	449	14	CD054203
	12	85.8	5.6	620	10	AW790714
	13	85	5.5	559	12	BM361770
	14	81.4	5.3	1800	9	AA415091
	15	75.2	4.9	781	29	CG811743
	16	67	4.4	875	12	BI952106
	17	63.6	4.2	931	12	BI488594
	18	62.2	4.1	751	14	CF342195
	19	61.4	4.0	604	14	CD056060
	20	61.2	4.0	780	12	BI753192
	21	61.2	4.0	902	14	CK249907
	22	60	3.9	611	13	BQ142729
	23	60	3.9	941	13	BQ225068
	24	60	3.9	1138	12	BM545507
	25	60	3.9	1194	14	CD014076
	26	60	3.9	1305	14	CD014075
	27	60	3.9	1364	14	CD014074
	28	60	3.9	1472	14	CD014072
	29	60	3.9	1496	14	CD014071
	30	60	3.9	1585	14	CD014073
	31	60	3.9	1723	29	AY407736
	32	60	3.9	1723	29	AY407737
	33	60	3.9	1745	11	BC036813
	34	60	3.9	2080	11	BC001541
	35	60	3.9	2187	11	BC026315
	36	59.6	3.9	657	14	CF523157
	37	59.6	3.9	671	14	CF524855
	38	59.6	3.9	681	14	CF524540
	39	59.6	3.9	685	14	CF523487
	40	59.6	3.9	698	14	CF343579
	41	59.6	3.9	722	14	CF525122
	42	59.6	3.9	723	14	CF524239
	43	59.6	3.9	724	14	CF525075
	44	59.6	3.9	755	14	CF523991
	45	59.6	3.9	757	14	CF374578

ALIGNMENTS

RESULT 1
CF881106/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CF881106 769 bp mRNA linear EST 31-OCT-2003
tric083xn18.b1 T.reesei mycelial culture, Version 6 October 2003
Hypocrea jecorina cDNA clone tric083xn18, mRNA sequence.

CF881106

EST.

Hypocrea jecorina (anamorph: Trichoderma reesei)

Hypocrea jecorina

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.

1 (bases 1 to 769)

Diener,S.E., Dankmeyer,L., Dunn-Coleman,N., Houfek,T.D.,

Mitchell,T.K., van Solingen,P., Teunissen,P.J.M., Ward,M. and

Dean,R.A.

Analysis of the protein processing and secretion pathways in a

Trichoderma reesei EST dataset

Unpublished (2003)

Contact: Ralph A. Dean

Fungal Genomics Laboratory

North Carolina State University

Campus Box 7251, Raleigh, NC 27695, USA

Tel: 919-513-0020

Fax: 919-513-0024

Email: ralph.dean@ncsu.edu

Seq primer: Lt-F1 primer.

FEATURES		Location/Qualifiers	
source		1. 769	
		/organism="Hypocrea jecorina"	
		/mol_type="mRNA"	
		/db_xref="taxon:51453"	
		/clone="tric083xn18"	
		/dev_stage="mycelia"	
		/clone_lib="T.reesei mycelial culture, Version 6 October 2003"	
		/note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."	
ORIGIN			
Query Match 11.3%; Score 173.2; DB 14; Length 769;			
Best Local Similarity 60.3%; Pred. No. 2.1e-20;			
Matches 364; Conservative 0; Mismatches 223; Indels 17; Gaps 4;			
QY	257	CGGTGCTTCCCGAGAGTGGAGTGGCTTCCATCAACATCACTGCTGGCGCGCGGGGACCA 316	
DB	663	CGGAGGACCAACAGTGGGAAGACTGCTTACATCTTCGAGCGGCTTTCCTCATTTGGCG 604	
QY	317	AGGCGGGGCGCAACCTCCGGTTCATGCTCTGGATCTTTGGCGGTGGGTTTGAGATCGGCA 376	
DB	603	AACCGGGGAGCGCTGCGGGTGGCTTTTACATCTTCGAGCGGCTTTCCTCATTTGGCG 544	
QY	377	GCCCCA-----CCATCTTCCCTCCCGCGCAGATGGTCAACAGAGTGTCTCATGGGCA 430	
DB	543	GGACGTGGCTGCTTCAACAGACCCCTCCAGCTGCTCCAGACGGGCGTTCGCTTCGGCA 484	
QY	431	AG-CCATCATCACTGCGCGTCACTACCGTGTGCTGCGGGGTTTCTGGCTGGT 489	
DB	483	AGCCCTTTCATCTTCGCGGGTCACTACCGGTGCGGGTGGGTTTCATGCTGGG 424	
QY	490	ATGACATCAAGGCGGAGGCGGGAACCGCGGCTTGAAGACACGAGCTTTGGGATGC 549	
DB	423	AGGAGATTCTCAATGAGGAGCGCAACCGCGGCTTGTTCACGAGCATGGGCTCG 364	
QY	550	AGTGGTGGGACAGCAATATGCGGGTTCGGGGGCGACCGAGAGTGTGATCATTTGG 609	
DB	363	AATGGTGGCGGATCAATGAGGCTTTGGCGGGAACCTCGCCGAGGTGACCATCTGG 304	
QY	610	CG-----AGCGGGGAGCATGTCGCTGTGTCGACCTCATCTGGAACGACGCGCAACA 665	
DB	303	GGAGTCCCTGGCTCCATTTCCGCTGGGACGAGTCTGTGTGTATGATGGGATGCA 244	
QY	666	CGTACAAGGGGCAAGCGGTTGTTCCGCGGGGCGATCATGCAAGGAGCCATGTTGCCG 721	
DB	243	CATACAACGACAAAGCTCTGTTCCGCGCGGCCATCATGAACCTCGGCTCCGTGACGCG 184	
QY	722	--GACCGGTGACGCGACGTACGCAACGAGATCTACGACCTTCTTGTCTCGAGTGTG 779	
DB	183	TCGATCCGCTGACTCGGAAAGGGGCAAGCGGCTGATCAGCATGTCGTCGAGAAGGCTG 124	
QY	780	GCTGTGGGAGCGGACGCAAGCTCGCGTGTGTTGGCAGTGGAGCGGACACCTTGCTCG 839	
DB	123	GATGGACGGGACGAGCAACCTTGGTGTCTCGGAACCTTGACCAACGAGCATTTGCG 64	
QY	840	ATGC 843	
DB	63	CCGC 60	
RESULT 2			
CB907899/c			
LOCUS			
DEFINITION			
tric083xn18 T.reesei mycelial culture, Version 3 april Hypocrea jecorina cDNA clone tric083xn18, mRNA sequence.			
ACCESSION			
CB907899			
VERSION			
CB907899.1 GI:30122557			
KEYWORDS			
Hypocrea jecorina (anamorph: Trichoderma reesei)			
SOURCE			
Hypocrea jecorina			
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.			
REFERENCE			
1 (bases 1 to 824)			
AUTHORS			
Foreman,P.K., Brown,D.E., Dankmeyer,L., Dean,R., Diener,S., Dunn-Coleman,N.S., Goedegebuur,F., Houfek,T.D., England,G.J., Kelley,A.S., Meerman,H.J., Mitchell,T., Mitchinson,C., Olivares,H.A., Teunissen,P.J., Yao,J. and Ward,M.			
TITLE			
Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus Trichoderma reesei			
J Biol. Chem. 278 (34), 31988-31997 (2003)			
JOURNAL			
MEDLINE			
PUBMED			
COMMENT			
22803314			
12788920			
Genecor Intl.			
Contact: Pamela K. Foreman			
925 Page Mill Road, Palo Alto, CA 94304, USA			
Tel: (650) 846-7635			
Fax: (650) 621-7817			
Email: Pforeman@genecor.com			
Seq primer: LT-F1 primer.			
Location/Qualifiers			
1. 824			
source			
/organism="Hypocrea jecorina"			
/mol_type="mRNA"			
/strain="QM6a"			
/db_xref="taxon:51453"			
/clone="tric083xn18"			
/dev_stage="mycelia"			
/clone_lib="T.reesei mycelial culture, Version 3 april"			
/note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."			
ORIGIN			
Query Match 11.3%; Score 173.2; DB 14; Length 824;			
Best Local Similarity 60.3%; Pred. No. 2.1e-20;			
Matches 364; Conservative 0; Mismatches 223; Indels 17; Gaps 4;			
QY	257	CGGTCTTCCCGAGAGTGGAGTGGCTTCCATCAACATCACTGCTGGCGCGCGGGGACCA 316	
DB	718	CGGAGGACCAACAGTGGGAAGACTGCTTACATCTTCGAGCGGCTTTCCTCATTTGGCG 659	
QY	317	AGGCGGGGCGCAACCTCCGGTTCATGCTCTGGATCTTTGGCGGTGGGTTTGAGATCGGCA 376	
DB	658	AACCGGGGAGCGCTGCGGGTGGCTTTTACATCTTCGAGCGGCTTTCCTCATTTGGCG 599	
QY	377	GCCCCA-----CCATCTTCCCTCCCGCGCAGTGGTCAACAGAGTGTCTCATGGGCA 430	
DB	598	GGAGTCCGCTGCTTCCCTTCAACGACCCCTCCAGTCTCCAGAGCGGCTTCGCTTCGGCA 539	
QY	431	AG-CCATCATCACTGCGCGTCACTACCGTGTGCTTCCCTGCTGGGGGTTTCTGGCTGGT 489	
DB	538	AGCCCTTTCATCTTCGCGGGTCACTACCGCTCGGGGTGGGGGTTTCATGCTGGGG 479	
QY	490	ATGACATCAAGCGCGGAGGCGGGAACCGCGGCTTGAAGACACGAGCTTTGGGATGC 549	
DB	478	AGGAGATTCTCAATGAGGAGCGCAACCGCGGCTTGTTCACGAGCCATGGGGCTCG 419	
QY	550	AGTGGTGGGACAGCAACATTGCGGGTTCGGCGGGACCCCGAGCAAGGTGACATCTTTGG 609	
DB	418	AATGGTGGCGCATCAACATTGAGGCTTTGGCGGGNACTCGTCCGAGGTGACCATCTGG 359	
QY	610	CG-----AGCGGGGAGCATGTCGCTGTTGTCGACCTCATCTTGGAAACGAGCGGCAACA 665	
DB	358	GGCAGTCCGCTGGCTCCATTTCGCTGCGGAGCAGCTCGTGTGTATGATGGCATGCA 299	
QY	666	CGTACAAGGGGCAAGCGGTTGTTCCGCGCGGCGATCATGCGAGGAGCCATGTTGCCG --- 721	
DB	298	CATACAACGACAAAGCTCTGTTCCGCGCGGCCCATCATGAACCTCGGGCTCCGTGACGCGCG 239	
QY	722	--GACCGGTGAGCGGACGAGTGTGCGGCAACGAGATCTTACGACCTCTTTGTCTCGAGTGTG 779	
DB	238	TCGATCCGCTCGACTCGGAAAGGGGCAAGCGGCTGTATCAGCATGTCGTGAGAAGGCTG 179	

QY 780 GCTGTGGCAGCGGACGACCAAGCTCGCGTGTGTGGCGAGTGGAGGACACCTTGTCTCG 839
 |||||
 Db 178 GATGCGAGCGGACGACGACACCTGGGTGTCTCCGGAACTTGACCAACGACGATTCG 119
 |||||
 QY 840 ATGC 843
 |||||
 Db 118 CCGC 115

RESULT 3
 CF880958
 LOCUS 719 bp mRNA linear EST 31-OCT-2003
 DEFINITION trico83xc07.b1 T.reesei mycelial culture, Version 6 October 2003
 Hypocrea jecorina cDNA clone trico83xc07, mRNA sequence.
 ACCESSION CF880958
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Hypocrea jecorina (anamorph: Trichoderma reesei)
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
 1 (bases 1 to 719)
 Diener S.E., Dankmeyer, L., Dunn-Coleman, N., Houfek, T.D.,
 Mitchell, T.K., van Solingen, P., Teunissen, P.J.M., Ward, M. and
 Dean, R.A.
 Analysis of the protein processing and secretion pathways in a
 Trichoderma reesei EST dataset
 Unpublished (2003)
 Contact: Ralph A. Dean
 Fungal Genomics Laboratory
 North Carolina State University
 Campus Box 7251, Raleigh, NC 27695, USA
 Tel: 919-513-0020
 Fax: 919-513-0024
 Email: ralph.dean@ncsu.edu
 Seq primer: IT-F1 primer:
 Location/Qualifiers
 1..719
 /organism="Hypocrea jecorina"
 /mol_type="mRNA"
 /strain="QM6a"
 /db_xref="taxon:51453"
 /clone="trico83xc07"
 /dev_stage="mycelia"
 /clone_lib="T.reesei mycelial culture, Version 6 October
 2003"
 /note="Vector: PREP3Y; Site 1: Not I/Sal I; Mycelial
 culture grown from 24 hrs to 6 days with varying Carbon
 and Nitrogen sources and concentrations."

ORIGIN
 Query Match 10.4%; Score 159.2; DB 14; Length 719;
 Best Local Similarity 61.4%; Pred. No. 5e-18;
 Matches 349; Conservative 0; Mismatches 198; Indels 21; Gaps 5;

QY 459 CGTGTGCTCGTGGGGTCTTGTGGTGGTGATGATCAAGCGCCGAGCGGACGGGGAAC 518
 |||||
 Db 1 CGCGTCGCGCGTCTTGTCTTCTGGCGCGCGGAGATCTCTGGCAACGGCAGCGGAAC 60
 |||||
 QY 519 GCGGCTTGAAGCAGCAGGCTTTGGCATGTCAGTGGTGGCAGACAACATTCGCGGTTC 578
 |||||
 Db 61 CTCGGCTGCTGACGAGCGCATGGCCCTGCGAGTGGTGGCGACAACATCGCGGCTTT 120
 |||||
 QY 579 GCGGCGACCCGACGAAAGTGA-CATCTTTGGCGAG--GCGGGCAGCATGTCGGTGTG 634
 |||||
 Db 121 GCGGCGACCCGACAAGGTGACCATCTGGGGCGAGTCGCGGGCGGCCATGTCGCTTAC 180
 |||||
 QY 635 TGCCACCTCATCTGGAACGACGCGCAACAGCTACAAAGGCGAAGCGGTGTTCGCGCGG 694
 |||||
 Db 181 GACCAGATGGTCTGTACGCGCGCAACGACGATACACGCGAAGCCCTGTTCGCGCG 240
 |||||
 QY 695 GGCATCATGCA-----GGGAGCCATGGTGTGCCGACCCGGTGGACGGCACTGACGGCAAC 748
 |||||

Db 241 GCCATCATGAACCTTGCAACGGCGTCCCGGAGCGGTGCACTCGCCACGCGCAG 300
 |||||
 QY 749 GAGATCTACACCTCTTTGTCTGAGTGTGGTGTGGTGTGGAGCGGCAAGCTCGCG 808
 |||||
 Db 301 GCCATCTACGACAACGTCGTGAAAAAGGCGCGGTGACAGCGCTCCGCGACTCGCTCGCG 360
 |||||
 QY 809 TGCTTGGCAGTGGAGCGACAC---TTGCTGATGCCACCAACACACTCTCTGGGTTC 865
 |||||
 Db 361 TGCTCGCGCAGTGGCTTACAGCAAGTTCCTCGCGCCCTCAACGTGGCGCGGCATC 420
 |||||
 QY 866 TTGGCGTACTCTCTGTGGCGTTGTACTTC-----CGGCCCGAGCGCAAGAACATCA 917
 |||||
 Db 421 CTGTCTTACAGCTCCCTGGGGGCTGTCTTACCTACCTGCGCGCCCGAGCGGTCTCTCTCC 480
 |||||
 QY 918 CCGATGACATGTACAAGTTGGTGGCGAGCGGCAAGTATGCAAGCGTTCCTGATCATTTG 977
 |||||
 Db 481 CGGACAGCCCCGAAAGCGCTCTCTGACGAGGCGAGTACCACGCGCTGCCCATGATTGCCG 540
 |||||
 QY 978 GCGACCAAGACGAGGAGGCGACCATCTT 1005
 |||||
 Db 541 GCGACGAGGAGGAGGCGACGCTCTT 568
 |||||

RESULT 4
 CB907713
 LOCUS 774 bp mRNA linear EST 02-JUL-2003
 DEFINITION trico83xc07 T.reesei mycelial culture, Version 3 april Hypocrea
 jecorina cDNA clone trico83xc07, mRNA sequence.
 ACCESSION CB907713
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Hypocrea jecorina (anamorph: Trichoderma reesei)
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
 1 (bases 1 to 774)
 Foreman, P.K., Brown, D.E., Dankmeyer, L., Dean, R., Diener, S.,
 Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J.,
 Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C.,
 Olivares, H.A., Teunissen, P.J., Yao, J. and Ward, M.
 Transcriptional regulation of biomass-degrading enzymes in the
 filamentous fungus Trichoderma reesei
 J. Biol. Chem. 278 (34), 31988-31997 (2003)
 22803314
 12788920
 Contact: Pamela K. Foreman
 Genencor Intl.
 925 Page Mill Road, Palo Alto, CA 94304, USA
 Tel: (650) 846-7635
 Fax: (650) 621-7817
 Email: Pforeman@genencor.com
 Seq primer: IT-F1 primer:
 Location/Qualifiers
 1..774
 /organism="Hypocrea jecorina"
 /mol_type="mRNA"
 /strain="QM6a"
 /db_xref="taxon:51453"
 /clone="trico83xc07"
 /dev_stage="mycelia"
 /clone_lib="T.reesei mycelial culture, Version 3 april"
 /note="Vector: PREP3Y; Site 1: Not I/Sal I; Mycelial
 culture grown from 24 hrs to 6 days with varying Carbon
 and Nitrogen sources and concentrations."

ORIGIN
 Query Match 10.4%; Score 159.2; DB 14; Length 774;
 Best Local Similarity 61.4%; Pred. No. 6e-18;
 Matches 349; Conservative 0; Mismatches 198; Indels 21; Gaps 5;

QY 459 CGTGTGCTCGTGGGGTCTTGTGGTGGTGATGATCAAGCGCGGAGCGGCGGAAC 518
 |||||
 Db 1 CGCGTCGCGCGTCTTGTCTTCTGGCGCGCGGAGATCTCTGGCAACGGCAGCGGAAC 518
 |||||

Fungal Genomics Laboratory
North Carolina State University
Campus Box 7251, Raleigh, NC 27695, USA
Tel: 919-513-0020
Fax: 919-513-0024
Email: ralph.dean@ncsu.edu

Seq primer: LT-F1 primer.
Location/Qualifiers

FEATURES
source

1..1070
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/clone="tric083xc05"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 6 October 2003"
/note="vector: pREP3Y; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN

Query Match 7.7%; Score 117.4; DB 14; Length 1070;
Best Local Similarity 70.0%; Pred. No. 1.6e-10;
Matches 187; Conservative 0; Mismatches 76; Indels 4; Gaps 2;
QY 459 CGTGTGCTCGTGGGGTTCTTGGCTGTGATGACATCAAGCGCGGAGCGGGAAC 518
Db 1 CGGTGCGCGGCTTGGCTTCTTGGCGGCGCGGAGATCTGCGCAAGCGGCGGAAC 60
QY 519 GCCGGTTGAAGACCAAGCTTTGGGATCGAGTGGTGGGAGACAACTGCGGGTTC 578
Db 61 CTGGGCTGTGACCAAGCGGCTGCGTGGGCGGCGGAGTGGGCGGCGGCTTT 120
QY 579 GCGGCGCAACCGAGCAAGTGA-CATCTTTGGCGAG--CGGCGACATGTCGTTG 634
Db 121 GCGGCGCAACCGAGCAAGTGA-CATCTTTGGCGAGTGGGCGGCGGCGGCTAC 180
QY 635 TGCCACTCATCTGGAACGCGGCGCAACACGCTACAAAGGCGGAGCGGTTGTCGGCGG 694
Db 181 GACCAAGTGTGTGTAGCGGCGCAAGCGGCGGCGGAGCGGCGGCGGCGGCGG 240
QY 695 GGCATCATGCGGAGCGATGTCGCG 721
Db 241 GCATCATGAATCTGCGCAGCGCGCG 267

RESULT 7

CF880590 708 bp mRNA linear EST 31-OCT-2003
LOCUS
DEFINITION
tric082xh02.b1 T.reesei mycelial culture, Version 6 October 2003
ACCESSION
VERSION
CF880590
CF880590.1 GI:38135272
KEYWORDS
SOURCE
EST.

Hypocrea jecorina (anamorph: Trichoderma reesei)
Hypocrea jecorina

REFERENCE
AUTHORS
1 (bases 1 to 708)
Hypocrea jecorina; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreales; Hypocrea; Hypocrea.

Diener, S.E., Dankmeyer, L., Dunn-Coleman, N., Houfek, T.D.,
Mitchell, I.K., van Solingen, P., Teunissen, P.J.M., Ward, M. and
Dean, R.A.

TITLE
Analysis of the protein processing and secretion pathways in a
Trichoderma reesei EST dataset

JOURNAL
MEDLINE
PUBMED
Unpublished (2003)

COMMENT
Contact: Ralph A. Dean
Fungal Genomics Laboratory
North Carolina State University
Campus Box 7251, Raleigh, NC 27695, USA
Tel: 919-513-0020
Fax: 919-513-0024
Email: ralph_dean@ncsu.edu

FEATURES

source

Seq primer: LT-F1 primer.
Location/Qualifiers
1..708
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/clone="tric082xh02"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 6 October 2003"
/note="vector: pREP3Y; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN

Query Match 7.2%; Score 110; DB 14; Length 708;
Best Local Similarity 60.3%; Pred. No. 2.7e-09;
Matches 216; Conservative 0; Mismatches 140; Indels 2; Gaps 2;
QY 249 GTTCCAGGCGGTCTCCCGAGTGAAGTGGCTTCCATCAACCTGAGTGGGCGGCC 308
Db 351 GTTCTGCAAAACCATGACCGACCAAGAAAGTGAATATCTGTCAGCGGCCCAA 410
QY 309 GGGCACCAAGCGCGGCGCCAACTCCCGGTCTGCTGATCTTTGGCGTGGTGTGA 368
Db 411 GGAGACCGAGCGGGGACAGTGGCTGTGACTTTGGATATACGGCGGACGCTTCCA 470
QY 369 GATCGGAGCGCCACCATCTTCCCTCCCGCCAGATGTTCCCAAGAGTGT-GCTCATGG 427
Db 471 GCTCGGGCGCCACCTCACGTACGACGGGACGAGCTCTATCTCGGACGTCGCCAGGG 530
QY 428 GCAAGCATCATCCAGTGGCGCTCACTACCGTGTGCTCGTGGGGTTCCTGGGTGG 487
Db 531 CCAGCCCTTCTATGAGGCGCTCACTACCGCTGCGACGCTTCTCTGGGCGG 590
QY 488 TGATGATCATCAAGCGCGGAGCGGAAACGCGGCTTCAAGGACCGAGCTTTGGGSCAT 547
Db 591 CGAGGAGATCTCGGCAACGGCAACGCGAACCTTGGGCTGCTGAACGACATGGGCT 650
QY 548 GCAAGTGGTGGAGACAATTCGCGGTTTGGCGGCGACCCGAGCAAGGTGACATCT 605
Db 651 GCAGTGGTGGCACAC-ACATCGAGGCTCTGCTGCTGCGACCGACAGGTGACCATCT 707

RESULT 8

CB907475 762 bp mRNA linear EST 02-JUL-2003
LOCUS
DEFINITION
tric082xh02 T.reesei mycelial culture, Version 3 april Hypocrea
jecorina cDNA clone tric082xh02, mRNA sequence.
ACCESSION
VERSION
CB907475
CB907475.1 GI:30122133
KEYWORDS
SOURCE
EST.

Hypocrea jecorina (anamorph: Trichoderma reesei)
Hypocrea jecorina

REFERENCE
AUTHORS
1 (bases 1 to 762)
Hypocrea jecorina; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreales; Hypocrea; Hypocrea.

Foreman, P.K., Brown, D.E., Dankmeyer, L., Dean, R., Diener, S.,
Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J.,
Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C.,
Olivares, H.A., Teunissen, P.J., Yao, J. and Ward, M.

TITLE
Transcriptional regulation of biomass-degrading enzymes in the
filamentous fungus Trichoderma reesei

JOURNAL
MEDLINE
PUBMED
J. Biol. Chem. 278 (34), 31988-31997 (2003)

COMMENT
Contact: Pamela K. Foreman

Genencor Intl.
925 Page Mill Road, Palo Alto, CA 94304, USA
Tel: (650) 846-7635
Fax: (650) 621-7817
Email: Pforeman@genencor.com

Seq primer: LT-F1 primer.

Location/Qualifiers

FEATURES

source

1. .762
/organism="Hypocrea jecorina"
/strain="QM6a"
/db_xref="taxon:51453"
/clone="trio82xh02"
/dev_stage="mycelia"
/note="Vector: pREP3; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN

Query Match 7.2%; Score 110; DB 14; Length 762;
Best Local Similarity 60.3%; Pred. No. 2.8e-09;
Matches 216; Conservative 0; Mismatches 140; Indels 2; Gaps 2;
QY 249 GTTCCAGCGGTGTTCCCGAGAGTGGAGTGGCTCAGCATCAACGTGGTGGCGCGCC 308
Db 405 GTTCTGCAAAACCATGACCGACCAAGAGAGTGAAGTGAACATATCTGTCAGCGGCCCAA 464
QY 309 GGGCACCAAGCGGGCGCCAACTCCCGTCTATGCTCTGGATCTTTGGCGGTGGGTTTGA 368
Db 465 GGAGACCAAGCGGGGGAAGTGGCTTGTGATATACGGCGGACGCTTCCA 524
QY 369 GATCGGACGCCCCACCATCTTCCCTCCGCGCAGATGTTACCAAGATGT-GCTCATGG 427
Db 525 GCTCGGGCCACCTCAGTACGAGCGGACAGCCCTCTATCTCGGACGTGGCCAGGG 584
QY 428 GGAAGCATATCAGTGGCGCTCACTACCGTGTGCTGCTGGGGTCTTCTGGCTGG 487
Db 585 CCAGCCCTTCTATCTAGAGCGGTCACTACCGGTGGACGGCTTCTGGCTTCTGGCGG 644
QY 488 TGATCATCAAGCGCGGAGCGGGAACCGCGCTTGAAGGACCAAGCGTTTGGGCAT 547
Db 645 CGACGAGATCTTGGCGACGGCAACGCGAACCTCGGCTGTGAACGACATGGGCT 704
QY 548 GCAGTGGGTGGCAGACAACTTCCCGGTGGCGGCGGACCGGAGAGGTGACATCT 605
Db 705 GCAGTGGGTGGCAGAC-ACATCGAGGCTCTGGTGGACACCGACAGGTGACCATCT 761

RESULT 9

CG810164

LOCUS

DEFINITION

clone KMFv5K21, genomic survey sequence.

CG810164.1

GI:38263638

GSS.

ORGANISM

Fusarium virguliforme

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

Hypocreomycetidae; Hypocreales; Nectriaceae; Nectria.

1 (bases 1 to 763)

Lightfoot,D.A. and Town,C.D.

End sequencing of BACs from a fingerprint physical map of the

causative agent of soybean sudden death syndrome, Fusarium

virguliforme (2003)

Unpublished

Other_GSSs: FSA071TR

Contact: Chris Town and K. Meksem

The Center of Excellence in Soybean Research, Teaching and

Outreach, Southern Illinois University at Carbondale and Plant

Genomics, The Institute for Genomic Research

Room 176, Ag. Building, Mail Code 4415, Carbondale, IL 62901-4415,

USA and 9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 618 453 3103 and 301-838-3523

Fax: 618 453-7457 and 301-838-0208

Email: meksemk@uiu.edu; cdtown@tigr.org (URL)

JOURNAL

COMMENT

http://Fusariumvirguliform.siu.edu)

Seq primer: TGTAAACGACGGCCAGT

Class: BAC ends.

FEATURES

source

Location/Qualifiers
1. .763
/organism="Fusarium virguliforme"
/mol_type="genomic DNA"
/cultivar="Monticello"
/db_xref="taxon:232082"
/clone="KMFv5K21"
/clone_lib="LargeInsertGenomicLibrary"
/note="Organ: Hyphae; Vector: pINDIGOBAC5; A single spore derived culture was used. Hyphae were grown in an incubator for four days. Nuclei were isolated and embedded in agarose, restriction digested with Hind III. Large size DNA fragments were ligated in vector pINDIGOBAC5 and electro-transformed into DH10B cells."

ORIGIN

Query Match 6.8%; Score 104.8; DB 29; Length 763;
Best Local Similarity 53.4%; Pred. No. 2.3e-08;
Matches 310; Conservative 0; Mismatches 262; Indels 8; Gaps 4;
QY 25 CGACACCATACCGGTCTCAACGCCATCATCAAGAGCGGTCTCTCGCATTCCTTTGC 84
Db 185 CGTACCGTCTATCGGCTCCAGCGGGCAAGGTGAATCATTAGGGGCATACCTTCGC 244
QY 85 CGAGCGCGGTGGGCAACCTCCGCTTCAAGAGCCCTGTGCGCTACTCTGGCTCGCTCAA 144
Db 245 AGACCCCTACTGTGTTCCCTACGCTCAACCTCCGAAGAACTCTCAAGAGCCCTGG 304
QY 145 CGGCGCAGAAAGTACTTACGCGCCGTGATGAGAGAACCCCGAGGGGACGTTTGAAG 204
Db 305 AAATTTTGA-TGCACTCTGCTCATCGGCCCATCATGCTCTCAATGTTCATCTCTAC 363
QY 205 AACCTTGGCAAGCGGAC-----TCGACTTTGGTATGCAGTGCAGTTCAGAGGT 259
Db 364 GGGCTGAGGACGTCACTCTCAAGTTCTCTGACTTTCTGAGCATCCCTTCTCTCCAG 423
QY 260 TGCTTTCCCGACAGTGCAGTGCCTCAGCATCAACGTGTGTGGCGCGCGGCGGACCA 319
Db 424 TTGTCACTGGCAAGAGACTGTCTCACCATGACTGTTTCAGCGTCTCTGTCGCCACCA 483
QY 320 CGGCGCGCAACCTCCCGGTCAFGCTCTGGATCTTTGGCGGTGGGTTTGAATCGGCG 379
Db 484 CTGGGCGACAAGCTTCTCTTTTGGATCTTTGGCGGTGGCTTCGAGCTCGGCACGA 543
QY 380 CCACCATCTTCTCTCCCGCCAGATGGTCAACA-AGAGTGTCTCATGGGCAAGCATCA 438
Db 544 GTGCCACGTATGATGGCACTAGCCTTCTGGGCGACAGGCATTCACAGGACCGAGCT 603
QY 439 TCCAGTGGCGGTCAACTACCGTGTTCCTCTGGGGGTTCCTTGGCTGGTGTGATCA 498
Db 604 TGTGTTGCTGTCACTATCGGTGTCTGGCTTGTGTTTCATGCTGGAGCGAGCTTC 663
QY 499 AGGCGGAGGCGAGCGGGAACCGCGCTTGAAGACAGCGGTTTGGGCATGCACTGGT 558
Db 664 AGAAAGA-GGCAGCACAAACCTAGTGTTCCTTGATCAACGAATAGACTGGAATGG 722
QY 559 CAGACACATTCGCGGTTCGGCGCGCGAGCCCGAGCAAGT 598
Db 723 CGGATAACATTGGCGCTTTTGGCGGTGACCCCTGAAAGGT 762

RESULT 10

CD054203

LOCUS

DEFINITION

HO02B22S HO Hordeum vulgare cDNA clone HO02B22 5-PRIME, mRNA

sequence.

CD054203

ACCESSION

VERSION

KEYWORDS

SOURCE

QY 450 GTCACTACCGTGTTCGCTGGGGTCTTCTGGCTGGTATGACATCAAGCGCGGGC 509
 |||||
 Db 250 GTCAACTATCGTCTTGGTCTTTTGGATGGCTCGCTGGAAGAGGTATTGGATGCTGGG 309
 |||||
 QY 510 AGCGGAACCGCGCTTGAAGGACAGCGTTTGGGCATGCGAGTGGTGGCGAGCAACATT 569
 |||||
 Db 310 CTTACAAATTGGGACACTACGATCAGATTCTTCTCTGCAATGGGTTCAAGACAACATT 369
 |||||
 QY 570 GCCGGTTCGGCGCGACCGCCAGCAAGGTGACATCTTTGG 609
 |||||
 Db 370 GCTAAGTTTCGGTGGAGACCGACAGCAAGGTTACTCTCTTTG 409
 |||||

RESULT 12
 AW790714 620 bp mRNA linear EST 01-MAY-2001
 LOCUS D00142-R Lambda Zap, Stratagene Blumeria graminis f. sp. hordei
 DEFINITION cDNA clone D00142 similar to lipase 2 precursor, mRNA sequence.

AW790714
 AW790714.1 GI:13902311

EST.

Blumeria graminis f. sp. hordei
 Blumeria graminis f. sp. hordei
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomyces;
 Erysiphales; Erysiphaceae; Blumeria.

1 (bases 1 to 620)
 Thomas, S.W., Rasmussen, S.W., Glaring, M.A., Rouster, J.A. and
 Oliver, R.P.

Gene identification in the fungal pathogen Blumeria graminis by
 expressed sequence tag analysis
 Unpublished (2000)

CONTACT: Rasmussen, S.W.

Department of Yeast Genetics

Carlsberg Laboratory

10 Gl. Carlsbergvej, DK-2500, Copenhagen, Denmark

Tel: 45 3327 5230

Fax: 45 3327 4766

Email: swr@crc.dk

High quality sequence stop: 620

POLYA=No.

FEATURES

source

1..620

/organism="Blumeria graminis f. sp. hordei"

/mol_type="mRNA"

/db_xref="taxon:62688"

/clone="D00142"

/cell_type="conidia"

/lab_host="Hordeum vulgare"

/clone_lib="Lambda Zap, Stratagene"

ORIGIN

Query Match 5.6%; Score 85.8; DB 10; Length 620;

Best Local Similarity 51.8%; Pred. No. 4.7e-05;

Matches 302; Conservative 0; Mismatches 262; Indels 19; Gaps 4;

QY 424 ATGGGCAAGCCATCATCCACGTGGCGTCAACTACCGTGTGCTCGTGGGGTCTTGG 483

Db 35 ACGGAAGCCAATTGTCTACATTGCGTCACTACCGTCTTGGTGGCTTGGATGTTAG 94

QY 484 CTGGTGATGACATCAAGCGCGGCGAGCGCGGAAACCGCGCTTGAAGCACCGCTTGG 543

Db 95 CTGGAAGAGAGGTGTGGATGCGCGCTTACCAACTTGGGACACTACGATCAGATTCTTG 154

QY 544 GATGCACTGGTGGGAGACACATTCGCGGTTCGGCGCGACCGGACCGAGGTGACAT 603

Db 155 CTCTACATGGGTTCAAGATAACATGCCAAATTTCGGTGGTATCCAGAAAAGTTACTC 214

QY 604 CTTTGGCGA---GGCGGCGAGCATGTCGGTGTGTGCCACCTCATCTGGAAGACGCGG 659

Db 215 TTTTCGGTACTCTGACAGGTGGTATCTCCACATGGAACCTGTTGACTGCCAAGACGGTA 274

QY 660 ACAACACGTACAAGGGCAACCGGTTGTTCCGCGCGGCGCATCATGCGAGGGACCATGGTG- 718

|||||

Db 275 ACCTTCCTCAACGACAAGCCTCTGTTCCGAGTCCATCATCGACAGTGAAGTGTTA 334
 |||||
 QY 719 -----CCGACCCCGTGGACGCGACGATCTACGACCTCTTTGTCTCGA 773
 |||||
 Db 335 TCCCAACTGACCTGCCGATTGTCCCAAGCCCAAAACATCTTCAACACTGTCTGTTCTGTG 394
 |||||
 QY 774 GTGCTGCTGTGGCAGCGCCAGCAGCAAGCTCGGCTGCTTGGCAGT---GCGAGCGACA 830
 |||||
 Db 395 CTGGAGGCTGTGAAGGAGCTGCTGACAAGATTGACTGCTCCGAGCTCTGCATACGACC 454
 |||||
 QY 831 CTTGCTCGATGCCACCAACAACACTCCTGGGTTCTTGGCGTACTCTCTCGTGTGGGTTGT 890
 |||||
 Db 455 AATTCTCCAAAGTAGTACATCTGTGCTGTGCTCTTTCGACTACCGTCTCTCGCTCTTT 514
 |||||
 QY 891 ACTCC-----CGCCCGGACGCGCAAGAACATCACCGATGACATGTACAAAGTTGGTGGCG 944
 |||||
 Db 515 CATACATTCACGATACGACGCTGATTGTTGACAGATTCTGCCGACAACTTTGTGAGG 574
 |||||
 QY 945 ACGCAAGATGCAAGGTTCCCGTGTATCATTTGGGACCAAA 987
 |||||
 Db 575 CTGGAAGATATGCCAAAGTCTCTTACATTTGCGGTGACCAAGAA 617
 |||||

RESULT 13

BM361770

LOCUS

DEFINITION

precursor, mRNA sequence.

BM361770

KEYWORDS

ORGANISM

Blumeria graminis f. sp. hordei

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomyces;

Erysiphales; Erysiphaceae; Blumeria.

1 (bases 1 to 559)

Glaring, M.A., Rasmussen, S.W., Oliver, R.P. and Thomas, S.W.

An expressed sequence tag analysis of the genes expressed during

appressorium formation in the barley mildew pathogen Blumeria

graminis

Unpublished (2002)

Contact: Rasmussen, S.W.

Department of Yeast Genetics

Carlsberg Laboratory

10 Gl. Carlsbergvej, DK-2500, Copenhagen, Denmark

Tel: 45 3327 5230

Fax: 45 3327 4766

Email: swr@crc.dk

POLYA=No.

FEATURES

Location/Qualifiers

1..559

/organism="Blumeria graminis f. sp. hordei"

/mol_type="mRNA"

/db_xref="taxon:62688"

/tissue_type="Appressorium forming conidia"

/clone_lib="Appressorium stage EST library of Blumeria

graminis f. sp. hordei"

/note="Vector: pBluescript II XR, Stratagene"

ORIGIN

Query Match 5.5%; Score 85; DB 12; Length 559;

Best Local Similarity 53.5%; Pred. No. 6.3e-05;

Matches 251; Conservative 0; Mismatches 205; Indels 13; Gaps 3;

QY 424 ATGGGCAAGCCATCATCCACGTGGCGTCAACTACCGTGTGCTCGTGGGGTCTTGG 483

Db 54 ACGGAAGCCAATTGTCTACATTGCGTCACTACCGTCTTGGTGGCTTGGATGTTAG 113

QY 484 CTGGTGATGACATCAAGCGCGGCGAGCGCGGAAACCGCGCTTGAAGGACCGCTTGG 543

Db 114 CTGGAAGGAGGTGTTGGATGCGCGGCTTACCAACTTGGGACACTACGATCAGATTCTTG 173

|||||

QY 544 GCATGCACTGGTGGCAGACAAATTCGCGGGTTCGGCGGACCCGAGCAAGGTGACAT 603
 Db 174 CTCACAAATGGTTCAGATAAATCGCAATTCGGTGGTGGTATCCAGAAAGGTACTC 233
 QY 604 CTTTGGGCGA-----GCGGGGAGCATGTCGGTGTGTCACCTCATCTGGAACGAGCG 659
 Db 234 TTTTCGGTACTTTCGAGGTGTAATCTCCACATGGAACCTGTTGACTGCCAAGACGGTA 293
 QY 660 ACAAACAGTCAAGGGCAAGCGTGTTCGCGGGGCGCATCATGACGAGGAGCCATGGT-- 717
 Db 294 ACCTTCGCTACAACGACAAGCGCTCTGTTCGAGCTGCCATCATGACAGTGGAGTGTGA 353
 QY 718 ----GCGGGACCGGTGACGCGACGTCAGCAACAGAGATCTACGACCTTTTGTCTGA 773
 Db 354 TCCCACTGACCCCTGCCGATGTGCCAAGGCCCAAAACATCTTCAACACTGCTGCTCGTG 413
 QY 774 GTGTGCTGTGGCAGCGCCAGCAAGCTCGCGTGTGCTGCGCAGT---GCGAGCGACA 830
 Db 414 CTGGAGGCTGTGAAGGAGCTGCTGACAAGATGACTGCTCCGAGCTTCGCCATACGACC 473
 QY 831 CTTTGTCTGATGCAACCAACACTCTCTGGGTTCTTGGGCTACTCCTC 879
 Db 474 AATTCTCTCAAGCTAGTACATCTGTGCTGTGCTCTCTCGACTACCCGTC 522

RESULT 14

AA415091 1800 bp mRNA linear EST 09-DEC-1999
 LOCUS Mg0026 RCW Lambda Zap Express Library Magnaporthe grisea cDNA clone
 DEFINITION RCW26 similar to Triacylglycerol Lipase (BC 3.1.1.3), mRNA
 SOURCE sequence.

AA415091
 VERSION AA415091.1 GI:2537256
 KEYWORDS EST.
 ORGANISM Magnaporthe grisea (anamorph: Pyricularia grisea)
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Sordariomycetes incertae sedis; Magnaportheaceae; Magnaporthe.

REFERENCE 1 (bases 1 to 1800)
 WU, S.-C., Bernstien, B.D., Darvall, A.G. and Albersheim, P.
 Expressed sequence tags of the rice blast fungus grown on rice cell
 walls
 Unpublished (1997)
 CONTACT: Sheng-Cheng Wu
 CCRC

University of Georgia
 220 Riverbend Road, Athens, GA 30602-4712, USA
 Tel: 706 542 4446
 Fax: 706 542 4412
 Email: wusc@bscr.uga.edu
 Fully sequenced
 Insert Length: 1800 Std Error: 0.00.

Location/Qualifiers

1. 1800
 /organism="Magnaporthe grisea"
 /mol_type="mRNA"
 /strain="CP987"
 /db_xref="taxon:148305"
 /clone="RCW26"
 /sex="Mat1-1 hermaphrodite"
 /tissue_type="mycelium"
 /dev_stage="Day 5 post-inoculation"
 /clone_lib="RCW Lambda Zap Express Library"
 /note="Vector: pBluescript excised from Lambda Zap
 Express; Site 1: EcoRI; Site 2: XhoI; Day 5
 post-inoculation MRNAs prepared from Magnaporthe grisea
 grown at 23C in the dark with constant gyratory shaking
 100 rpm in Vogel's minimal medium containing 0.5% isolated
 rice cell walls as the sole carbon source. Library
 provided by Sheng-Cheng Wu. Sequences were processed by
 one of two methods. Where a full-length alignment to the
 M. grisea genome sequence was available, the EST sequence
 was trimmed according to the alignment, otherwise sequence

FEATURES

source

quality was assessed using phredPhrap version 991019 and
 trimmed according to phd files (0.05) and for vector
 seqs."

ORIGIN

Query Match 5.3%; Score 81.4; DB 9; Length 1800;
 Best Local Similarity 49.7%; Pred. No. 0.0041;
 Matches 373; Conservative 0; Mismatches 341; Indels 37; Gaps 5;
 QY 274 GAGACTGCTCACCATCAACGTTGGTGGCGCGCCGCGGACCAAGGGGGCGCCAAACCTC 333
 Db 408 GAAGACTGTCTTTTCTCTCGACATCTACGCGCGTCCAAATGCCACTTTCATCTCCAAAGTC 467
 QY 334 CCGTTCATGCTCTGATCTTTGGCGGTGGTGTGAGATCGCAGCCCCACCATCTTCCT 393
 Db 468 CCGTGTGTCAATTCATCCAGGGCGGGGTTCACAAAATTCCAAATCCAACTCCAACTACGAC 527
 QY 394 CCGCCCCAGATGTTCAACAAGAGTGTCTCATGGCAAGCATCATCCAGTGGCCCTCA 453
 Db 528 GGCACAGGGCTGGTCAAG-----GCCGGGATTAAGGCTGATGCCATATCTTCA 579
 QY 454 ACTACCGTGTGCTCTGTTGGGGTCTTGTGATGACATCAAGGCCGAGGCGACG 513
 Db 580 ACTACCGCTCGGGCCCTACGGTCTCTCAC-----CAACGGCAACGAGTCTGAGC 630
 QY 514 GGAACCGCGGCTTGAAGGACCGGTTTGGCATGCGAGTGGTGGGACAACTTCCCG 573
 Db 631 CCAACACGGGTAAAGGACCGCAAGCGCTCCAGTGGATCAAGAAGAACAATTTCGCG 690
 QY 574 GGTTCGGCGGCGACCCGAGCAAGGTGACATCTTTGGCGAGGGCGGAGCAGTGTCCGTGT 633
 Db 691 AGTTCCGTGGCGACCG-----GACCACGTCGTCCTGATCGCGGTTTCGCGCGAGG 742
 QY 634 GTGCCACCTCATCTGGAAACGACGGCGACAAACAGTCAAGGGCAAGCGTGTTCGCGCG 693
 Db 743 GCGGACATCTCCCTGCTGATCGGTACGCGCGCAAGGAGGAGGCGCTGTTCACGCG 802
 QY 694 GGCATCATGCGAGGAGCCATGTCGCGGACCCGCGTGGAGCGGACGTCAGGCAACGAGAT 753
 Db 803 CGCAGCGCGCACCGCGGTGTGCTGCGACGCGTGTGACCGTGGAGGAGTCCCACTACCA 862
 QY 754 CTACGACCTCTTTGTCTCGAGTGTGCTGTCAGCGCGCAGCGCAAGCTCCGCGTCTT 813
 Db 863 GTACGACAACTTTGCCATCTCGAGGCTGTGGGTGCG-----CGACTCGTGGCTGCTT 919
 QY 814 GGCAGTGGCAGGACACCTTGTGTCGATGCCAACCAACACTCTCTGGGTCTTTCGCGTA 873
 Db 920 GCGGTCCAAAAACACCCACGAGCTGACGCGCGCCCAACAGCGGCTTACCGGGGCG 979
 QY 874 CTCCTCGTTGCGGTGTACTCCCGGCG-----GAGGCAAGAACTACCGATGA 924
 Db 980 CTCGAGGCGCGCTGTGTATGTGAACCCCGTGTGACGCGCATCTGACGAGCT 1039
 QY 925 CATGTACAAGTTGTGTCGCGCGCAAGTATGCAAGCGTTCCTCGTATCATTTGGCGACCA 984
 Db 1040 GACCTATACCGCTTCGAGGAGGAGGTTCTTCGCGGTTCGCGTCACTTCGCGACGTC 1099
 QY 985 GAACGACGAGGCAACATCTTTGGTCTTTGA 1015
 Db 1100 AACCAACGACGGCGGTCTTTGCCCTCGA 1130

RESULT 15

CG811743/3
 LOCUS CG811743/3
 DEFINITION FSAAS27TF LargeInsertGenomicLibrary Fusarium virguliforme genomic
 clone KMVF5F5, genomic survey sequence.
 ACCESSION CG811743
 VERSION CG811743.1 GI:38265217
 KEYWORDS GSS.
 SOURCE Fusarium virguliforme
 ORGANISM Fusarium virguliforme
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

